CONFIDENCE SETS IN BOUNDARY AND SET ESTIMATION

HANNA K. JANKOWSKI*, LARISSA I. STANBERRY

Abstract. Let $p_1 \leq p_2$ and consider estimating a fixed set $\{x : p_1 \leq f(x) \leq p_2\}$ by the random set $\{x : p_1 \leq \widehat{f}_n(x) \leq p_2\}$, where $\widehat{f}_n$ is a consistent estimator of the continuous function f. This paper gives consistency conditions for these sets, and provides a new method to construct confidence regions from empirical averages of sets. The method can also be used to construct confidence regions for sets of the form $\{x : f(x) \leq p\}$ and $\{x : f(x) = p\}$. We then apply this approach to set and boundary estimation. We describe conditions for strong consistency for the empirical average sets and study the fluctuations of these via confidence regions. We illustrate the proposed methods on several examples.

1. Introduction

The ability to estimate a set and its boundary is of primary importance in many fields. As an example, consider estimating the domain of covariates which yield dangerous blood pressure levels.

In this paper, we study the estimation of sets of the form

$$
\begin{aligned}
F(p) &= \{x \in \mathcal{D} : f(x) \leq p\}, \\
\partial F(p) &= \{x \in \mathcal{D} : f(x) = p\}, \qquad (1.1)\\
F(p_1, p_2) &= \{x \in \mathcal{D} : p_1 \leq f(x) \leq p_2\},
\end{aligned}
$$

where $\mathcal{D} \subset \mathbb{R}^d$ is a compact set, and f is a continuous function, $f : \mathcal{D} \mapsto \mathbb{R}$. Note that the boundary of the set $F(p)$ may not necessarily be equal to $\partial F(p)$, but we use the notation nonetheless. Let $\widehat{f}_n$ denote a consistent estimator of f. We estimate the sets (1.1) using "plug-in" estimators obtained by replacing f with $\widehat{f}_n$ in the definition:

$$
\begin{aligned}
\widehat{F}_n(p) &= \{x \in \mathcal{D} : \widehat{f}_n(x) \leq p\}, \\
\partial \widehat{F}_n(p) &= \{x \in \mathcal{D} : \widehat{f}_n(x) = p\}, \qquad (1.2)\\
\widehat{F}_n(p_1, p_2) &= \{x \in \mathcal{D} : p_1 \leq \widehat{f}_n(x) \leq p_2\}.
\end{aligned}
$$

Our first goal is to provide conditions on the consistency of these estimators. Next, to assess the accuracy of the estimators, we construct appropriate confidence regions: random sets which cover the sets (1.1) with at least a $100(1-\alpha)\%$ probability. We

Date: March 10, 2009.

Key words and phrases. random closed set, oriented distance functions, set expectation, boundary estimation, level set, simultaneous confidence interval.

*Corresponding author.

refer to these as confidence sets or confidence regions to differentiate from confidence intervals or simultaneous confidence bands in function estimation.

Sets of the form (1.1) appear in various statistical problems, such as estimation of contour clusters [Nol, Pol] or the estimation of density support [KT]. Here, we consider the problem of estimating the domain of covariates with specified response level(s), which arise for example in treatment comparisons [TWS] and toxic dose estimation [BT].

Another application of the proposed method is set and boundary estimation, and a large part of this paper is dedicated to this topic. The observed data here are random sets: primate skulls, mammograms, or cell images. Population characteristics may be summarized by the expected set (or expected boundary), which in turn are estimated using empirical averages. In [SJ], new definitions of the expectation of a random set and the expected random boundary were given. The definitions have a number of desirable properties, particularly suitable for problems in image analysis. Moreover, both the expected set and expected boundary have the same form as in (1.1). We apply our results to determine conditions for consistency of the empirical averages, and study when and if these conditions fail under a variety of random set models. Furthermore, we construct confidence regions for the expected set and the expected boundary. Thus, jointly with [SJ], our work provides a statistical framework for inference about random sets and their boundaries.

Suppose then that $\widehat{f}_n$ is a random, continuous function such that

$$\text{(A1)} \qquad \sup_{x\in\mathcal{D}} |\widehat{f}_n(x) - f(x)| \to 0 \text{ almost surely as } n \to \infty.$$

The sets (1.1) are estimated by (1.2). Consistency of $\widehat{F}_n(p)$ as an estimator for $F(p)$ was studied in [Mol]. We extend these results to the sets $\partial\widehat{F}_n(p)$ and $\widehat{F}_n(p_1, p_2)$. We then show how to construct simultaneous confidence sets for these estimators, under the assumption of weak convergence

$$\text{(A2)} \qquad \sqrt{n}\{\widehat{f}_n(\cdot) - f(\cdot)\} \Rightarrow \mathbb{Z}(\cdot),$$

where $\mathbb{Z}(\cdot)$ is a continuous random field on $\mathcal{D}$. The methods extend naturally to the case where the scaling required is not $\sqrt{n}$, however, we restrict our discussion to this setting. We expect that, for the $\sqrt{n}$ scaling, the limiting field $\mathbb{Z}$ will most often be Gaussian, or a Gaussian transform. For increased accuracy, the confidence sets may be restricted to any compact window $\mathcal{W} \subset \mathcal{D} \subset \mathbb{R}^d$. Although fluctuation results exist for $F(p)$ and for mean sets (see eg. [Cre, BM, Mol]), these appear to be of a theoretical quality. To our knowledge, the work presented here is the first attempt to construct confidence regions for level sets and expected random sets.

The outline of this paper is as follows. Section 2 looks at the consistency of the sets (1.2), and Section 3 discusses how to form the confidence regions. Sections 3.1 and 4 cover the two main applications of our methods: covariate domain estimation, and

random set and boundary estimation. In Section 5, we give some further examples. Proofs of all results appear in the Appendix.

1.1. **Notation and Assumptions.** Unless otherwise stated, we assume that $\mathcal{D}$ is the working domain and write, for example, $F(p) = \{x : f(x) \leq p\}$ without stating that $x \in \mathcal{D}$ explicitly. As previously noted, we assume that $\mathcal{D}$ is a compact subset of $\mathbb{R}^d$, and denote the Euclidean norm of x as $|x|$. In practice, $\mathcal{D}$ will most likely be a simple geometric shape, such as a d-dimensional rectangle, or a d-dimensional ball.

We write $B_r(x_0) = \{x : |x - x_0| \leq r\}$ for the closed ball of radius r centred at x_0. For a set A, we write $A^o, \overline{A}, A^c$ and ∂A to denote its interior, closure, complement and boundary. Unless noted otherwise, set operations are calculated relative to the domain $\mathcal{D}$. That is, $A^c = \mathcal{D} \setminus A$, and so forth. Furthermore, for a set A, we define $A^\delta = \{x : B_\delta(x) \cap A \neq \emptyset\} = \cup_{x \in A} B_\delta(x)$. Deterministic sets are denoted using capital letters $A, B \ldots$, while bold upper-case lettering, $\boldsymbol{A}, \boldsymbol{B}, \ldots$, is used for random sets.

The notation $C(\mathcal{D})$ is used to denote the space of continuous functions $C(\mathcal{D}) = \{f : \mathcal{D} \mapsto \mathbb{R},\ f \text{ continuous}\}$ endowed with the uniform metric. We write $X_n \Rightarrow X$ to say that X_n converges weakly to X. Throughout the paper, when handling weak convergence of stochastic processes or random fields, we assume that they take values in $C(\mathcal{D})$.

2. Consistency

Let $\mathcal{F}$ be the family of closed sets of $\mathbb{R}^d$ and let $\mathcal{K}$ denote the family of all compact subsets of $\mathbb{R}^d$. For a probability triple $(\Omega, \mathcal{A}, P)$, a random closed set is the mapping $\boldsymbol{A} : \Omega \mapsto \mathcal{F}$ such that for every compact set $K \in \mathcal{K}$

$$\{\omega : \boldsymbol{A}(\omega) \cap K \neq \emptyset\} \in \mathcal{A},$$

(cf. [Mat]). We write r.c.s. for random closed set, although the notation RACS is also used in the literature. Note that

$$\begin{aligned}
\{\widehat{F}_n(p) \cap K \neq \emptyset\} &= \left\{\inf_{x \in K} \widehat{f}_n(x) \leq p\right\}, \\
\{\widehat{F}_n(p_1, p_2) \cap K \neq \emptyset\} &= \left\{\inf_{x \in K} \left|\widehat{f}_n(x) - \frac{p_1 + p_2}{2}\right| \leq \frac{p_2 - p_1}{2}\right\}, \\
\{\partial\widehat{F}_n(p) \cap K \neq \emptyset\} &= \left\{\inf_{x \in K} |\widehat{f}_n(x) - p\,| \leq 0\right\}.
\end{aligned}$$

Therefore, since by assumption the functions $\widehat{f}_n$ are continuous almost surely, the estimators (1.2) satisfy the measurability requirement and are well-defined.

Recall that the Hausdorff distance between two sets, A and B, is defined as

$$\rho(A, B) = \inf\left\{\delta > 0 : A \subset B^\delta, B \subset A^\delta\right\}.$$

Following [Mol], we say that a r.c.s. $\boldsymbol{A}_n$ *converges strongly* to a deterministic set A if $\rho(\boldsymbol{A}_n, A) \to 0$ almost surely. For other notions of convergence for random sets see, for example, [SVW, SW].

The key conditions for the consistency of the estimators (1.2) are

$$\{x : f(x) \leq p\} = \overline{\{x : f(x) < p\}} \tag{2.3}$$

$$\{x : p \leq f(x)\} = \overline{\{x : p < f(x)\}}. \tag{2.4}$$

Proposition 2.1. *Condition* (2.3) *is equivalent to* $\partial\{x : f(x) \geq p\} = \{x : f(x) = p\}$, *and condition* (2.4) *is equivalent to* $\partial\{x : f(x) \leq p\} = \{x : f(x) = p\}$.

Theorem 2.2 (Theorem 2.1 in [Mol])**.** *Under assumption (A1), the estimator* $\widehat{F}_n(p)$ *converges strongly to* $F(p)$ *if the condition* (2.3) *holds at* p.

Note that this theorem also holds if f and $\widehat{f}_n$ are lower semi-continuous. We now extend the results to $\widehat{F}_n(p_1, p_2)$ and $\partial\widehat{F}_n(p)$.

Theorem 2.3. *Under assumption (A1), the estimator* $\widehat{F}_n(p_1, p_2)$ *converges strongly to* $F(p_1, p_2)$ *if the function* f *satisfies condition* (2.3) *at* $p = p_2$ *and condition* (2.4) *at* $p = p_1$. *Moreover,* (2.3) *and* (2.4) *are necessary in the following sense:*

1. *Suppose that* x_0 *is a point such that there exists a neighbourhood* $B_\delta(x_0)$ *and a subsequence* n_k *such that* $\widehat{f}_{n_k}(x) > f(x)$ *for all* $x \in B_\delta(x_0)$. *If* $\widehat{F}_n(p_1, p_2)$ *is consistent, then* (2.3) *must hold at* x_0 *for* $p = p_2$ *in the sense that*
$$x_0 \notin \{x : f(x) \leq p_2\} \setminus \overline{\{x : f(x) < p_2\}}.$$
2. *Suppose that* x_0 *is a point such that there exists a neighbourhood* $B_\delta(x_0)$ *and a subsequence* n_k *such that* $\widehat{f}_{n_k}(x) < f(x)$ *for all* $x \in B_\delta(x_0)$. *If* $\widehat{F}_n(p_1, p_2)$ *is consistent, then* (2.4) *must hold at* x_0 *for* $p = p_1$ *in the sense that*
$$x_0 \notin \{x : p_1 \leq f(x)\} \setminus \overline{\{x : p_1 < f(x)\}}.$$

Corollary 2.4. *Under assumption (A1), the estimator* $\partial\widehat{F}_n(p)$ *converges strongly to* $\partial F(p)$ *if the function* f *satisfies both* (2.3) *and* (2.4) *at* p.

The conditions (2.3) and (2.4) arise in three different ways: through local minima, local maxima, and through flat regions. For example, suppose that $\partial F(p) = \{x : f(x) = p\}$ is a connected subset of $\mathcal{D} \subset \mathbb{R}^d$ with positive Lebesgue measure in $\mathbb{R}^d$. Then neither (2.3) nor (2.4) hold at p, see Figure 1 (left). On the other hand, if $f(x)$ has a local minimum on $\partial F(p)$, then (2.3) does not hold (Figure 1, middle), whereas if $f(x)$ has a local maximum on $\partial F(p)$, then (2.4) does not hold (Figure 1, right).

The examples in Figure 1 also clarify the necessity of the conditions. Suppose that f is of the form shown in Figure 1 (right) and we want to estimate the set $\{x : 0 \leq f(x) \leq 0.2\} = \{0.5\}$. If a subsequence $\{n_k\}$ of $\widehat{f}_n(x)$ approaches f from below, then for this subsequence $\widehat{F}_{n_k}(0, 0.2) = \emptyset$ and consistency does not hold.

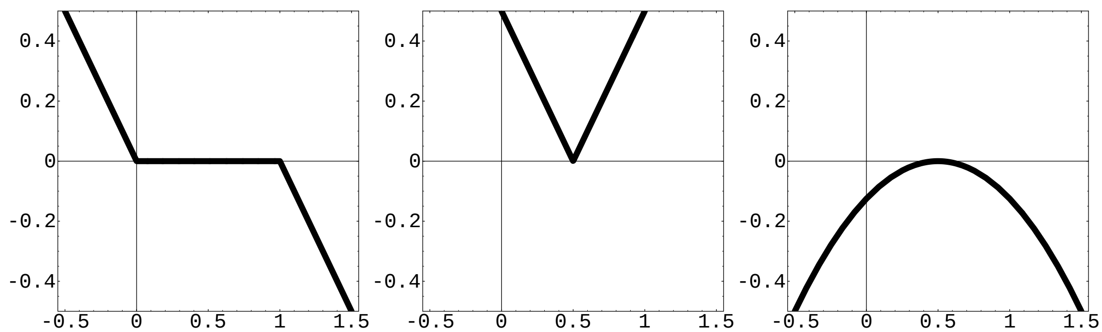

FIGURE 1. Examples of functions which do not satisfy the consistency conditions for $p = p_1 = p_2 = 0$: (right) the function does not satisfy (2.4); (centre) the function does not satisfy (2.3); (left) the function satisfies neither (2.3) nor (2.4). Here, $\mathcal{D} = [-0.5, 1.5] \subset \mathbb{R}$.

The following proposition shows that increasing functions always satisfy the consistency conditions.

Proposition 2.5. *Suppose that f is continuous and that there exists a direction $e_0 \in \mathbb{R}^d$ such that $f(x)$ is strictly increasing in the direction e_0 for all $x \in \mathcal{D}$. That is, suppose that there exists an e_0 such that for all $x_0 \in \mathcal{D}$, the function $h(t) = f(x_0 + te_0)$ is strictly increasing as a function of $t \in \mathbb{R}_+$. Then f satisfies conditions* (2.3) *and* (2.4) *for any value of p.*

Remark 2.6. *Note that if f is differentiable, then the condition of the previous proposition is satisfied if, for some $e_0 \in \mathbb{R}^d$,*

$$\nabla f(x) \cdot e_0 > 0 \quad \textit{for all } x \in \mathcal{D}.$$

EXAMPLE 1 (disc with random centre). Let $\mathcal{D} = [-2, 2]^2 \subset \mathbb{R}^2$ with $f(x) = |x|$ and $F(1) = \{x : f(x) \leq 1\}$, i.e. $F(1)$ is the disc with radius one centred at the origin. Suppose $U_1, \ldots, U_n$ are IID random variables from the uniform distribution on $[-1, 1]^2$, and let $\bar{U}_n$ denote their bivariate sample mean. Then $\widehat{f}_n(x) = |x - \bar{U}_n|$ converges uniformly to $f(x)$ on $\mathcal{D}$, and $f(x)$ satisfies (2.3) and (2.4) at $p = 1$. Therefore, $\widehat{F}_n(1)$ and $\partial\widehat{F}_n(1)$ are consistent for $F(1)$ and $\partial F(1) = \{x : |x| = 1\}$. In this case, the Hausdorff distance $\rho(\widehat{F}_n(1), F(1)) = \rho(\partial\widehat{F}_n(1), \partial F(1)) = |\bar{U}_n|$ converges to zero almost surely. Figure 2 shows an example of the estimator for several values of n.

Remark 2.7. *We note that in both Theorem 2.2 and Theorem 2.3 the restriction that $\mathcal{D}$ is compact may be removed. For a closed and unbounded domain $\mathcal{D}$, assumption*

(A1) for convergence of $\widehat{f}_n$ should be replaced with

$$\sup_{x \in K \cap \mathcal{D}} |\widehat{f}_n(x) - f(x)| \to 0 \quad \textit{as } n \to \infty \textit{ almost surely,}$$

for all compact sets $K \subset \mathbb{R}^d$.

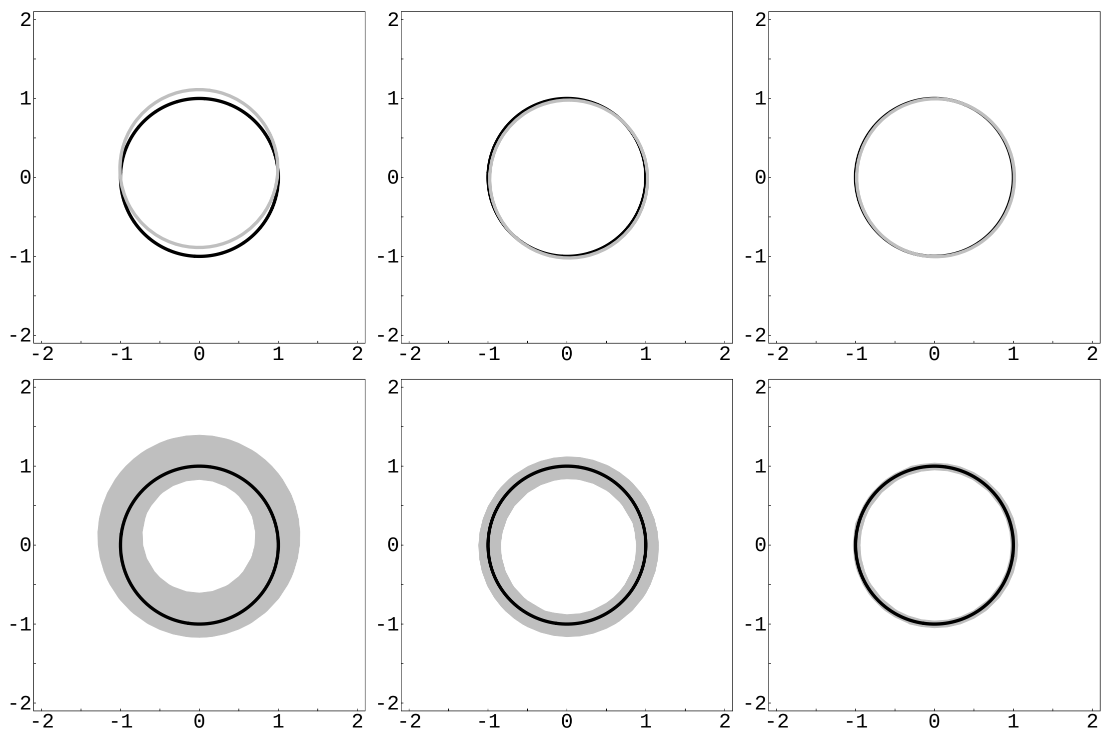

Figure 2. For each of $n = 25, 100, 1000$ (from left to right), the estimate $\partial\widehat{F}(1)$ (Example 1) is shown in grey in the top row, with its 95% confidence interval (Example 2) shown in the bottom row. The true set $\partial F(1)$ is shown in black for comparison.

3. Confidence Regions for Level Sets

Suppose that the estimating functions $\widehat{f}_n$ satisfy assumption (A2). That is, the empirical fluctuation field $\{\mathbb{Z}_n(x)\}_{x\in\mathcal{D}}$ converges to the continuous field $\{\mathbb{Z}(x)\}_{x\in\mathcal{D}}$ weakly in $C(\mathcal{D})$:

$$\mathbb{Z}_n(\cdot) \equiv \sqrt{n}(\widehat{f}_n(\cdot) - f(\cdot)) \Rightarrow \mathbb{Z}(\cdot).$$

Let $\mathcal{W} \subseteq \mathcal{D} \subset \mathbb{R}^d$ be a compact set and define q_1 and q_2 to be the quantiles of the process $\sup_{x\in\mathcal{W}} \mathbb{Z}(x)$ such that

$$\mathrm{pr}\left(\sup_{x\in\mathcal{W}} \mathbb{Z}(x) \leq q_1\right) = 1-\alpha \quad \text{and} \quad \mathrm{pr}\left(\sup_{x\in\mathcal{W}} |\mathbb{Z}(x)| \leq q_2\right) = 1-\alpha.$$

Then

$$\begin{aligned} &\left\{x \in \mathcal{W} : \widehat{f}_n(x) \leq p + \tfrac{1}{\sqrt{n}}\, q_1\right\}, \\ &\left\{x \in \mathcal{W} : p_1 - \tfrac{1}{\sqrt{n}}\, q_2 \leq \widehat{f}_n(x) \leq p_2 + \tfrac{1}{\sqrt{n}}\, q_2\right\}, \\ &\left\{x \in \mathcal{W} : |\widehat{f}_n(x) - p| \leq \tfrac{1}{\sqrt{n}}\, q_2\right\} \end{aligned} \tag{3.5}$$

form $100(1-\alpha)\%$ confidence regions for the sets

$$\begin{aligned} F_{\mathcal{W}}(p) &\equiv \{x \in \mathcal{W} : f(x) \leq p\}, \\ F_{\mathcal{W}}(p_1, p_2) &\equiv \{x \in \mathcal{W} : p_1 \leq f(x) \leq p_2\}, \\ \partial F_{\mathcal{W}}(p) &\equiv \{x \in \mathcal{W} : f(x) = p\} \end{aligned}$$

respectively. If $\mathcal{W} = \mathcal{D}$ then $F_{\mathcal{W}}(p) = F(p)$ and so on. Note that the random variables $\sup_{x\in\mathcal{W}} |\mathbb{Z}(x)|$ and $\sup_{x\in\mathcal{W}} \mathbb{Z}(x)$ are well-defined because $\mathcal{W}$ is compact and $\mathbb{Z}$ has continuous sample paths. This also implies that we may use $\max_{x\in\mathcal{W}} |\mathbb{Z}(x)|$ and $\max_{x\in\mathcal{W}} \mathbb{Z}(x)$ to calculate the quantiles, which is computationally easier.

The confidence sets are conservative in the sense that they capture the set of interest at least $100(1-\alpha)\%$ of the time. To illustrate this, we consider estimating $\partial F(p)$, with $\alpha = 0.05$. The other cases follow a similar reasoning.

By definition of q_2, we know that with a probability of 95%,

$$\sup_{x\in\mathcal{D}} |\widehat{f}_n(x) - f(x)| \leq q_2/\sqrt{n},$$

and, if this holds, then the set of interest $\partial F(p)$ is contained within the set

$$\left\{x : p - \frac{q_2}{\sqrt{n}} \leq \widehat{f}_n(x) \leq p + \frac{q_2}{\sqrt{n}}\right\}.$$

Hence, the confidence set captures $\partial F(p)$ with at least 95% certainty. Alternatively, the probability that there exists an x which is not in the confidence region but satisfies $f(x) = p$ is at most 5%. The exact probability of missing a point of $\partial F(p)$ is given by $\mathrm{pr}\left(\sup_{x\in\partial F(p)} |\mathbb{Z}(x)| > q_2\right)$.

On the other hand, we want to make sure that points x such that $f(x) \neq p$, are not included in the confidence set for $\partial F(p)$. For a fixed ε, consider the set $\{x : |f(x) - p| \geq \varepsilon\}$. The probability that these are included in the confidence set for

$\partial F(p)$ is

$$\begin{aligned}
&\mathrm{pr}\left(\sup_{x:|f(x)-p|\geq\varepsilon} |\widehat{f}_n(x) - p| \leq \frac{q_2}{\sqrt{n}}\right) \\
\approx\ &\mathrm{pr}\left(\sup_{x:|f(x)-p|\geq\varepsilon} |\mathbb{Z}(x) + \sqrt{n}(f(x)-p)| \leq q_2\right).
\end{aligned}$$

This quantity behaves in a manner similar to type II error: for fixed ε it converges to zero as $n \to \infty$, and for fixed n it decreases as ε increases.

We also make the following remarks regarding the definition of the confidence set.

1. In practice, the quantiles q_1 and q_2 may not be straightforward to calculate exactly. However, they can be estimated using sampling or re-sampling methods such as the bootstrap. In this case, it may be easier to use the asymmetric quantiles $\mathrm{pr}(q_{21} \leq \min_{x\in\mathcal{D}} \mathbb{Z}(x)) = 0.025$ and $\mathrm{pr}(\max_{x\in\mathcal{D}} \mathbb{Z}(x) \leq q_{22}) = 0.975$ instead of q_2. This avoids the maximization of $|\mathbb{Z}(x)|$, which is often computationally more intensive. For asymmetric quantiles, the confidence sets for $F_{\mathcal{W}}(p_1, p_2)$ and $\partial F_{\mathcal{W}}(p)$ become
$$\left\{x \in \mathcal{W} : p_1 + \frac{q_{21}}{\sqrt{n}} \leq \widehat{f}_n(x) \leq p_2 + \frac{q_{22}}{\sqrt{n}}\right\},$$
and
$$\left\{x \in \mathcal{W} : \frac{q_{21}}{\sqrt{n}} \leq \widehat{f}_n(x) - p \leq \frac{q_{22}}{\sqrt{n}}\right\},$$
respectively.

2. Notice that the consistency conditions play no role in the design of the confidence sets (3.5). Indeed, the confidence region functions as intended even if consistency is violated. Example 14 in Section 4 illustrates this point.

3. The smoothness and variability of the field $\mathbb{Z}$ determines the "size" of the confidence set, which may not be uniform over $\mathcal{D}$. For this reason, we can select the set $\mathcal{W}$ to be a strict subset of $\mathcal{D}$, to obtain confidence regions for $\mathcal{W} \cap F(p) = F_{\mathcal{W}}(p)$, $\mathcal{W} \cap \partial F(p) = \partial F_{\mathcal{W}}(p)$ and $\mathcal{W} \cap F(p_1, p_2) = F_{\mathcal{W}}(p_1, p_2)$. Note that the larger the window $\mathcal{W}$ is chosen, the wider the confidence set is. Consider, for example, the estimation of $F(p)$ and suppose that $\mathcal{W}_1 \subset \mathcal{W}_2 \subset \mathcal{D}$. Let q_1, q_1^* be the quantiles obtained for the two confidence sets: $\mathrm{pr}(\sup_{x\in\mathcal{W}_1} \mathbb{Z}(x) \leq q_1) = \mathrm{pr}(\sup_{x\in\mathcal{W}_2} \mathbb{Z}(x) \leq q_1^*) = 0.95$. Then
$$\{x \in \mathcal{W}_1 : \widehat{f}_n(x) \leq q_1/\sqrt{n}\} \subseteq \mathcal{W}_1 \cap \{x \in \mathcal{W}_2 : \widehat{f}_n(x) \leq q_1^*/\sqrt{n}\}.$$
This idea is illustrated in Example 16.

4. In practice, assumption (A2) can be checked using the techniques described in [Bil, Var] for $\mathcal{D} \subset \mathbb{R}$ or [Kun] for $\mathcal{D} \subset \mathbb{R}^d$.

EXAMPLE 2 (confidence sets for Example 1). Using the methods outlined above, we derive a 95% confidence set for $\partial F(p) = \{x : |x| = 1\}$ with $\mathcal{W} = \mathcal{D} = [-2, 2]^2$. We calculate

$$\mathbb{Z}_n(x) = \sqrt{n}\left(|x - \bar{U}_n| - |x|\right),$$

where $\bar{U}_n$ is the average of n independent Uniform$[-1, 1]^2$ random variables. Clearly, $\mathbb{Z}_n(x)$ has continuous sample paths. Also, $|\mathbb{Z}_n(x)| \leq \sqrt{n}|\bar{U}_n|$ for all x and, since this is realized at $x = 0$, we obtain that $\sup_{x\in\mathcal{D}} |\mathbb{Z}_n(x)| = \sqrt{n}|\bar{U}_n|$. The limiting distribution is therefore easy to calculate exactly. Let $\{U_1^1, \ldots, U_n^1, U_1^2, \ldots, U_n^2\}$ be a sequence of independent Uniform$[-1, 1]$ random variables, and let

$$\bar{U}_n^i = \frac{1}{n}\sum_{j=1}^{n} U_j^i \quad , i = 1, 2.$$

Then

$$\sqrt{n}|\bar{U}_n| = \sqrt{(\sqrt{n}\,\bar{U}_n^1)^2 + (\sqrt{n}\,\bar{U}_n^2)^2} \Rightarrow \sqrt{3^{-1}(Z_1^2 + Z_2^2)},$$

where $\bar{U}_n = (\bar{U}_n^1, \bar{U}_n^2)$ and Z_1, Z_2 are independent standard normal variables. Hence,

$$95\% = \mathrm{pr}\left(\sup_{x\in\mathcal{D}} |\mathbb{Z}(x)| \leq q_2\right) = \mathrm{pr}\left(Z_1^2 + Z_2^2 \leq 3q_2^2\right)$$

and therefore $q_2 = \sqrt{5.99/3} = 1.41$. Figure 2 shows the resulting confidence set for $\partial F(1)$, for different values of n.

3.1. **Covariate Domain Estimation.** Consider a linear model of the form

$$f(x) = E[Y|x] = \beta \cdot \widetilde{x},$$

where Y is the observed response variable, x is the vector of covariates, $\beta = (\beta_0, \ldots, \beta_p)$ is the vector of parameters, and $\widetilde{x}$ is a function of the covariates. For example, if $f(x) = E[Y|x] = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_1 x_2$, then $d = 2, p = 3$ and $\widetilde{x} = (1, x_1, x_2, x_1 x_2)$. We stray from conventional notation to differentiate between the covariates measured (e.g. weight, blood pressure) and their use in the model (e.g. the constant or any cross-terms). Thus, $\widetilde{x} : \mathbb{R}^d \mapsto \mathbb{R}^{p+1}$ is a function of x and we assume that it is continuous. We also assume that the covariates are not categorical and lie within a compact set $\mathcal{D} \subset \mathbb{R}^d$.

Let $F(\alpha_0) = \{x \in \mathcal{D} : E[Y|x] \leq \alpha_0\}$ and $F(\alpha_0, \alpha_1) = \{x \in \mathcal{D} : \alpha_0 \leq E[Y|x] \leq \alpha_1\}$ denote the domain of covariates for which the response variable Y lies within the target range. We next use the proposed method to estimate $F(\alpha_0)$ and $F(\alpha_0, \alpha_1)$, and to construct confidence regions for these sets. Although we focus on the linear model, our methods can be easily extended to generalized linear models.

Recall that in the normal linear model, the maximum likelihood estimators $\widehat{\beta}_n$ of β are consistent, efficient and asymptotically normal. Consider the estimating equation

$$\widehat{f}_n(x) = \widehat{y} = \widehat{\beta} \cdot \widetilde{x},$$

where n is the number of observations. Since $\widetilde{x}$ is continuous, the image of $\mathcal{D}$ under $\widetilde{x}$ is compact. It follows that $\widehat{f}_n(x)$ converges uniformly to $f(x)$ and therefore, $\widehat{f}_n$ satisfies assumption (A1). The conditions (2.3) and (2.4) need to be checked on a case by case basis. For example, for $f(x) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_1 x_2$, both conditions are satisfied for any value of p, as long as at least one of $\beta_1, \beta_2, \beta_3$ is non-negative; this follows from Proposition 2.5.

Next, let $\mathbb{Z}_n(x) = \sqrt{n}(\widehat{f}_n(x) - f(x))$. As $\widehat{\beta}$ is asymptotically normal, we have that $\sqrt{n}(\widehat{\beta} - \beta) \Rightarrow Z$, where Z is multivariate normal with mean zero and variance Σ. If unknown, Σ is estimated using standard methods. Since $\widetilde{x}$ is continuous, it follows that $\mathbb{Z}_n(x)$ converges weakly in $C(\mathcal{D})$ to a continuous, mean zero Gaussian field, $\mathbb{Z}(x) = Z \cdot \widetilde{x}$, with covariance structure given by

$$\text{cov}(\mathbb{Z}(x), \mathbb{Z}(x')) \;=\; \widetilde{x}\,\Sigma \cdot \widetilde{x}.$$

Therefore, $\widehat{f}_n$ satisfies assumption (A2), and confidence sets may be formed as described above. Since $\mathbb{Z}(x)$ is linear in Z, approximations for the quantiles of $\sup_{x\in\mathcal{D}} \mathbb{Z}(x)$ and $\sup_{x\in\mathcal{D}} |\mathbb{Z}(x)|$ are calculated using either sampling or re-sampling methods (such as the parametric or nonparametric bootstrap). Depending on the form of $\widetilde{x}$ as a function of x, simplifications to $\sup_{x\in\mathcal{D}} \mathbb{Z}(x)$ are also possible, as shown in the next example.

Example 3. We illustrate the method on the data set `trees` available from [R]. Here, girth (in inches), height (in feet) and volume (in cubic feet) of timber were recorded for 31 felled black cherry trees. Set $x_1 =$ girth and $x_2 =$ height. Fitting the model

$$E[\log Y|x] = \beta_0 + \beta_1 \log x_1 + \beta_2 x_2,$$

we obtain estimates $\widehat{\beta}_0 = -6.63$ (p-value $= 5.1e-09$), $\widehat{\beta}_1 = 1.98$ (p-value $< 2e-16$) and $\widehat{\beta}_2 = 1.12$ (p-value $= 7.8e-06$). The estimates are not far from the formula volume $=$ height $\times$ girth$^2/4\pi$.

Set $\mathcal{D} = [5, 25] \times [50, 100]$, and suppose that we are interested in the domain of covariates for which the log-volume is at least $\log 30$ (≈ 3.4), that is,

$$\begin{aligned} F(-\log 30) &= \{x \in \mathcal{D} : E[\log Y|x] \geq \log 30\} \\ &= \{x \in \mathcal{D} : -\beta_0 - \beta_1 \log(x_1) - \beta_2 \log(x_2) \leq -\log 30\}. \end{aligned}$$

Figure 3 shows the estimator $\widehat{F}_n(-\log 30) = \{x \in \mathcal{D} : \widehat{f}_n(x) \leq -\log 30\}$, where

$$\widehat{f}_n(x) = -\widehat{\beta}_0 - \widehat{\beta}_1 \log(x_1) - \widehat{\beta}_2 \log(x_2) = \widehat{\beta} \cdot \widetilde{x},$$

and $\widetilde{x} = (-1, -\log(x_1), -\log(x_2))$. Note that $\widetilde{x}$ is continuous on $\mathcal{D}$.

The true function $f(x) = -\beta_0 - \beta_1 \log(x_1) - \beta_2 \log(x_2)$ is strictly decreasing in both x_1 and x_2. Therefore, by Proposition 2.5, it satisfies condition (2.3) at $p = -\log 30$, or for any other choice of p. Condition (2.4) is also satisfied, although we do not require it here. It follows that the set $\widehat{F}_n(-\log 30)$ is consistent for $F(-\log 30)$.

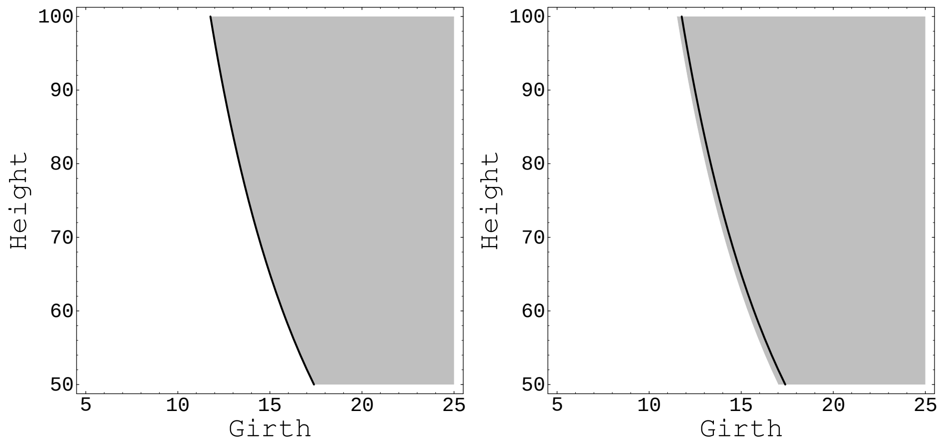

FIGURE 3. The estimated set $\widehat{F}_n(-\log 30)$ (left), with the 95 % confidence region (right). The boundary of the estimated set is superimposed in black for comparison.

The 95% confidence set for $F(-\log 30)$ is $\{x : \widehat{f}_n(x) \leq -\log 30 + q_1/\sqrt{31}\}$, where q_1 is the value such that $\mathrm{pr}(\sup_{x\in\mathcal{D}} \mathbb{Z}(x) \leq q_1) = 0.95$. In this case, the fluctuation process is $\mathbb{Z}(x) = Z \cdot \widetilde{x}$, where $Z = (Z_0, Z_1, Z_2)$ is multivariate Gaussian. Under the normal linear model, a consistent estimator of the covariance matrix of Z is

$$\begin{aligned}\widehat{\Sigma} &= \widehat{\sigma}^2(X'X)^{-1} \\ &= \begin{bmatrix} 0.6397 & 0.0208 & -0.1601 \\ 0.0208 & 0.0056 & -0.0081 \\ -0.1601 & -0.0081 & 0.0418 \end{bmatrix},\end{aligned}$$

where X is the design matrix of the regression. The supremum of $\mathbb{Z}$ must occur on one of the corners of $\mathcal{D}$,

$$\sup_{z\in\mathcal{D}} \mathbb{Z}(x) = \max\left\{\mathbb{Z}(a_1, b_1), \mathbb{Z}(a_1, b_2), \mathbb{Z}(a_2, b_1), \mathbb{Z}(a_2, b_2)\right\},$$

for $a_1 = \log 5, a_2 = \log 25, b_1 = \log 50$ and $b_2 = \log 100$. We estimate the quantile by repeated sampling from a multivariate Gaussian with variance $\widehat{\Sigma}$, to obtain

$$\{x \in \mathcal{D} : -\beta_0 - \beta_1 \log(x_1) - \beta_2 \log(x_2) \leq -\log 30 + 0.2303/\sqrt{31}\}$$

as the approximate 95% confidence region. The resulting set (Figure 3, right) is quite tight.

Note that for domain $\mathcal{D} = [0, 25] \times [50, 100]$, the function $\widetilde{x}$ is not continuous on $\mathcal{D}$, and hence our theory does not apply. Specifically, the function $\mathbb{Z}(x)$ is not bounded for this choice of $\mathcal{D}$, and confidence sets cannot be computed.

4. The Expected Set and Expected Boundary

In classical statistics, the object of interest takes values in $\mathbb{R}^d$, a linear space. This leads to a natural definition of the average and expectation. However, the family of closed sets is nonlinear, and therefore the expected set is not easily defined.

In [SJ], a new definition of the expectation and the expected boundary of a random closed set was given. The proposed definitions are easy to implement and have a number of desirable properties. In particular, they satisfy certain natural inclusion, equivariance and preservation properties. In what follows, we study the consistency of the boundary estimator and discuss the construction of confidence sets for the expected set and the expected boundary.

4.1. **Definition.** The definitions of the expected set and expected boundary are based on the oriented distance function (ODF), which has been studied extensively in the context of shape analysis [DZ2, DZ1].

Suppose that we observe data within a compact window $\mathcal{D} \subset \mathbb{R}^d$. The distance function is defined for any nonempty set $A \subset \mathcal{D}$ as

$$d_A(x) \;=\; \inf_{y \in A} |x - y| \quad \text{for } x \in \mathcal{D}.$$

We note that $d_A(x) \equiv d_B(x)$ if and only if $\overline{A} = \overline{B}$ and hence the distance function partitions the family of sets into equivalence classes of sets with equal closure. Given the distance function of A, the original set may be recovered (up to equivalence class) via $A = \{x : d_A(x) = 0\}$. Also, the Hausdorff distance may be calculated using the distance function as

$$\begin{aligned} \rho(A, B) &= \max\left\{\sup_{x \in A} d(x, B), \sup_{x \in B} d(x, A)\right\} \\ &= \sup_{x \in \mathcal{D}} |d_A(x) - d_B(x)|, \end{aligned}$$

for any sets $A, B \subset \mathcal{D}$ [DZ1].

The oriented distance function of any set $A \subset \mathcal{D}$ such that $\partial A \neq 0$ is given by

$$b_A(x) \;=\; d_A(x) - d_{A^c}(x) \quad \text{for } x \in \mathcal{D}.$$

The ODF takes into account both the set and its interior, and therefore provides more information than the distance function. Here, $b_A(x) \equiv b_B(x)$ if and only if $\overline{A} = \overline{B}$ and $\partial A = \partial B$, giving a finer partition of the family of sets. The set and its boundary may now be recovered (again, up to equivalence class) by $A = \{x : b_A(x) \leq 0\}$ and $\partial A = \{x : b_A(x) = 0\}$. Note that it is not necessary that the domain $\mathcal{D}$ be a compact set in the definition of the ODF. We make this assumption as it is necessary to compute confidence regions, and because it is a natural assumption to make in practice. Figure 4 shows some examples of oriented distance functions.

We also note here that there exist several efficient algorithms to calculate the distance function, and hence the ODF, of any set A [BGKW, FBF, RP]. This allows for easy implementation of our methods. More details are given in [SJ].

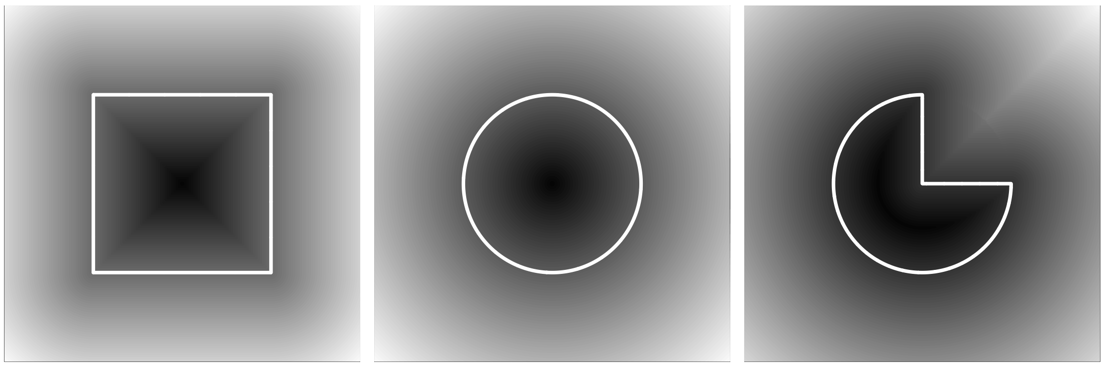

FIGURE 4. Examples of ODFs for different sets: a "filled" square (left), a disc (centre), and a "pacman" - a disc with the upper-right segment removed (right). The boundary of the set (white) is superimposed on the grey scale image of the ODF.

For a random closed set $\boldsymbol{A}$ we define the random function $b_{\boldsymbol{A}}(x)$, and denote its pointwise mean as $E[b_{\boldsymbol{A}}(x)]$. The mean set and mean boundary are then defined as follows.

Definition 4.1. *Suppose that $\boldsymbol{A}$ is a random closed set such that $\partial \boldsymbol{A} \neq \emptyset$ almost surely and assume that $E[|b_{\boldsymbol{A}}(x_0)|] < \infty$ for some $x_0 \in \mathcal{D}$. Then*

$$\begin{aligned} E[\boldsymbol{A}] &= \{x \in \mathcal{D} : E[b_{\boldsymbol{A}}(x)] \leq 0\}, \\ E[\partial \boldsymbol{A}] &= \{x \in \mathcal{D} : E[b_{\boldsymbol{A}}(x)] = 0\}. \end{aligned}$$

To provide some insight into this definition, we consider the following examples.

EXAMPLE 4 (disc with random radius). Suppose that $\boldsymbol{A} = \{x : |x| \leq \Theta\}$, for some integrable real-valued random variable Θ. Then $b_{\boldsymbol{A}}(x) = |x| - \Theta$ and hence $E[\boldsymbol{A}] = \{x : |x| \leq E[\Theta]\}$ and $E[\partial \boldsymbol{A}] = \{x : |x| = E[\Theta]\}$. That is, the expected set is a disc with radius $E[\Theta]$.

EXAMPLE 5 (interval with random centre). Let Θ be a Uniform$[-1, 1]$ random variable, and let $\boldsymbol{A} \subset \mathbb{R}$ be the r.c.s. given by $\{x : |x - \Theta| \leq 1\}$. We may think of this as a special case of the disc in $\mathbb{R}$ with random centre, in contrast to Example 1. Then

$$E[b_{\boldsymbol{A}}(x)] = \begin{cases} -x - 1, & x \leq -1 \\ 0.5x^2 - 0.5, & x \in [-1, 1] \\ x - 1, & x \geq 1 \end{cases},$$

and therefore $E[\boldsymbol{A}] = [-1, 1]$ and $E[\partial \boldsymbol{A}] = \{-1, 1\}$. Another example of a disc with random centre is also considered in Example 13.

4.2. **Consistency.** Suppose that we observe $\boldsymbol{A}_1, \ldots, \boldsymbol{A}_n$ random sets, and let

$$\bar{b}_n(x) = \frac{1}{n} \sum_{i=1}^{n} b_{\boldsymbol{A}_i}(x).$$

Then the empirical mean set and the empirical mean boundary are defined as

$$\bar{\boldsymbol{A}}_n = \{x : \bar{b}_n(x) \leq 0\} \quad \text{and} \quad \partial \bar{\boldsymbol{A}}_n = \{x : \bar{b}_n(x) = 0\}.$$

Consistency of these estimators may now be established using the results of Section 2.

Corollary 4.2. *Suppose that*

$$\lim_n \sup_{x \in \mathcal{D}} |\bar{b}_n(x) - E[b_{\boldsymbol{A}}(x)]| = 0 \quad \textit{almost surely,} \tag{4.6}$$

and that $E[\boldsymbol{A}]$ is well-defined. Suppose also that the expected ODF, $E[b_{\boldsymbol{A}}(x)]$, satisfies

$$\{x : E[b_{\boldsymbol{A}}(x)] \leq 0\} \quad = \quad \overline{\{x : E[b_{\boldsymbol{A}}(x)] < 0\}}. \tag{4.7}$$

Then $\bar{\boldsymbol{A}}_m$ converges strongly to $E[\boldsymbol{A}]$. If $E[b_{\boldsymbol{A}}(x)]$ also satisfies

$$\{x : E[b_{\boldsymbol{A}}(x)] \geq 0\} \quad = \quad \overline{\{x : E[b_{\boldsymbol{A}}(x)] > 0\}}, \tag{4.8}$$

then $\partial \bar{\boldsymbol{A}}_m$ converges strongly to $E[\partial \boldsymbol{A}]$.

Recall that Figure 1 explains these conditions for a general function f, and in this case we have that $f(x) = E[b_{\boldsymbol{A}}(x)]$ and $p = 0$. It is of interest to understand how the conditions (4.7) and (4.8) arise, if at all, for $E[b_A(x)]$ under different types of random set models. We explore this question by considering a variety of different examples.

Remark 4.3. *Condition* (4.6) *is of course condition (A1) for the random set ODFs. We say that the random sets $\boldsymbol{A}_1, \ldots, \boldsymbol{A}_n$ are independent and identically distributed (IID) if their oriented distance functions $b_{\boldsymbol{A}_1}, \ldots, b_{\boldsymbol{A}_n}$ are IID as random functions on $\mathcal{D}$. Under this assumption, condition* (4.6) *follows immediately. A more detailed discussion of independent random closed sets is given in the Appendix, Section 6.3.*

Example 6 (half plane). For $\mathcal{D} \subset \mathbb{R}^d$, consider the r.c.s. $\boldsymbol{A} = \boldsymbol{A}(\Theta) = \{x \in \mathcal{D} : x_1 \leq \Theta\}$, where Θ is a real-valued random variable with finite mean $E[\Theta]$. Then $b_{\boldsymbol{A}}(x) = x_1 - \Theta$, and $E[b_{\boldsymbol{A}}(x)] = x_1 - E[\Theta]$. The mean ODF satisfies both conditions (4.7) and (4.8), and therefore $\bar{\boldsymbol{A}}_n = \{x : x_1 \leq \bar{\Theta}_n\}$ and $\partial \boldsymbol{A}_n = \{x : x_1 = \bar{\Theta}_n\}$ are consistent estimators of $E[\boldsymbol{A}] = \{x : x_1 \leq E[\Theta]\}$ and $E[\partial \boldsymbol{A}] = \{x : x_1 = E[\Theta]\}$. Indeed, we may easily check that $\rho(E[\boldsymbol{A}], \bar{\boldsymbol{A}}_n) = \rho(E[\partial \boldsymbol{A}], \partial \boldsymbol{A}_n) = |\bar{\Theta}_n - E[\Theta]|$ which converges to zero almost surely.

Example 7 (set and its boundary). Suppose that $\boldsymbol{A} \subset \mathbb{R}$ is either $[0, 1]$ or $\{0, 1\}$ with equal probability. Then $E[\boldsymbol{A}] = E[\partial \boldsymbol{A}] = [0, 1]$. On the other hand, if $[0, 1]$ is seen with probability p, then

$$\text{if } p < 0.5, \quad E[\boldsymbol{A}] = E[\partial \boldsymbol{A}] = \{0,1\};$$
$$\text{if } p > 0.5, \quad E[\boldsymbol{A}] = [0,1] \text{ and } E[\partial \boldsymbol{A}] = \{0,1\}.$$

The case $p = 0.5$ provides a setting where neither (4.7) nor (4.8) are satisfied.

Suppose we observe n independent sets $\boldsymbol{A}_1, \boldsymbol{A}_2, \ldots, \boldsymbol{A}_n$, where each $\boldsymbol{A}_i$ is either $[0,1]$ or $\{0,1\}$ with equal probability. Let $\hat{p}_n$ denote the proportion of the random sets equal to $[0,1]$. Then

$$\bar{b}_n(x) = \hat{p}_n b_{[0,1]}(x) + (1-\hat{p}_n) b_{\{0,1\}}(x),$$

and it follows that whenever $\hat{p}_n < 0.5$, $\bar{\boldsymbol{A}}_n = \partial \bar{\boldsymbol{A}}_n = \{0,1\}$, and for $\hat{p}_n > 0.5$ $\bar{\boldsymbol{A}}_n = [0,1]$ while $\partial \bar{\boldsymbol{A}}_n = \{0,1\}$. Clearly, convergence to the expected set $[0,1]$ can never be achieved.

We may also consider empirical versions of the consistency conditions:

$$\begin{aligned} \{x : \bar{b}_n(x) \le 0\} &= \overline{\{x : \bar{b}_n(x) < 0\}}, \\ \{x : \bar{b}_n(x) \ge 0\} &= \overline{\{x : \bar{b}_n(x) > 0\}}. \end{aligned}$$

Returning to Example 7 with $\hat{p}_n < 0.5$, we calculate

$$\{0,1\} = \{x : \bar{b}_n(x) \le 0\} \neq \overline{\{x : \bar{b}_n(x) < 0\}} = \emptyset.$$

On the other hand, if $\hat{p}_n > 0.5$, we obtain that

$$[0,1] = \{x : \bar{b}_n(x) \le 0\} = \overline{\{x : \bar{b}_n(x) < 0\}} = [0,1].$$

Therefore, the empirical versions of (4.7) and (4.8) are not sufficient to ascertain convergence.

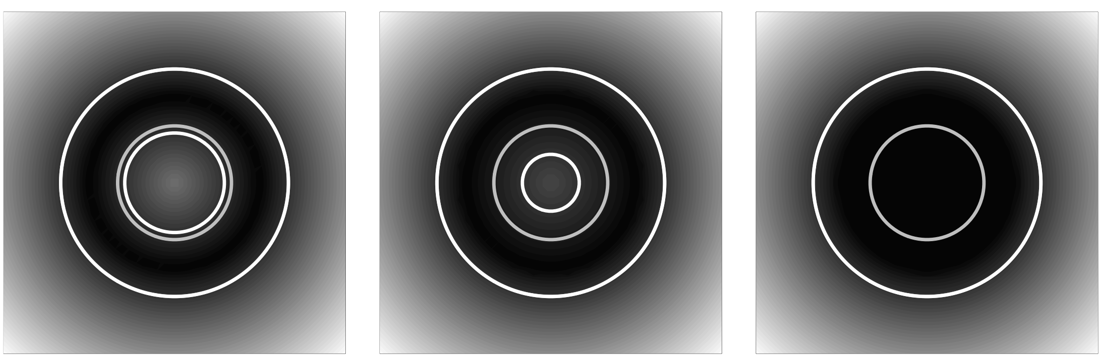

FIGURE 5. Mean sets for Example 8: The expected boundary (white) superimposed on a grey scale image of $E[b_{\boldsymbol{A}}(x)]$. The interior of the boundary (with darker values of the grey scale image) is the expected set. The three pictures correspond to $p = 0.1$ (left), $p = 0.25$ (centre), and $p = 1/2$ (right). For comparison, the boundary of the donut is also shown in grey.

Remark 4.4. *It is possible that $E[b_{\boldsymbol{A}}(x)]$ violates the conditions* (4.7) *and/or* (4.8) *and consistency still holds. For example, consider the r.c.s. $\boldsymbol{A} = \{x_0\} \subset \mathbb{R}^d$ almost surely. Then IID sampling trivially produces a consistent estimate, but $E[b_{\boldsymbol{A}}(x)]$ fails to satisfy* (4.7).

Although it fails to satisfy the consistency conditions, it is natural to view Example 7 as somewhat pathological and unrealistic. The next two examples are designed to study more realistic versions of removing the "middle" of a set.

EXAMPLE 8 (missing timbit). Suppose that $\boldsymbol{A}$ is either a disc or a donut in $\mathbb{R}^2$; that is,

$$\boldsymbol{A} = \begin{cases} \{x : |x| \leq 1\} & \text{with probability } p, \\ \{x : 0.5 \leq |x| \leq 1\} & \text{otherwise.} \end{cases}$$

Then the expected set $E[\boldsymbol{A}]$ is a donut for $p < 1/3$, and a disc for $p \geq 1/3$ (Figure 5). For $p \neq 1/3$, the expected ODF satisfies both (4.7) and (4.8). For $p = 1/3$, we have

$$E[b_A(x)] = \begin{cases} |x| - 1 & \text{for } |x| \geq 0.75, \\ -|x|/3 & \text{otherwise,} \end{cases}$$

and hence $E[\boldsymbol{A}] = \{x : |x| \leq 1\}$ while $E[\partial \boldsymbol{A}] = \{x : |x| = 0, 1\} \neq \partial E[\boldsymbol{A}]$. Since $E[b_{\boldsymbol{A}}(x)]$ has a local maximum at $x = 0$, it fails to satisfy (4.8), and therefore this point may be omitted by estimators $\partial \bar{\boldsymbol{A}}_n$.

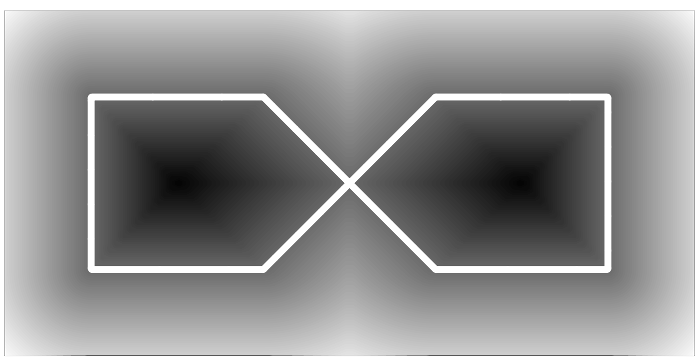

FIGURE 6. The expected boundary (white) for Example 9 is superimposed on a grey scale image of the expected ODF.

EXAMPLE 9 (blinking square). Suppose that the r.c.s. $\boldsymbol{A}$ is either a rectangle or a union of two squares with equal probability. Specifically, define

$$\begin{aligned} A_1 &= \{x : 0 \leq x_1 \leq 3, 0 \leq x_2 \leq 1\}, \\ A_2 &= \{x : 0 \leq x_1 \leq 1, 0 \leq x_2 \leq 1\} \cup \{x : 2 \leq x_1 \leq 3, 0 \leq x_2 \leq 1\}. \end{aligned}$$

Then $\boldsymbol{A} = A_1$ with probability 0.5, and otherwise $\boldsymbol{A} = A_2$. Thus, half of the time the $\boldsymbol{A}$ has its "middle" removed. The resulting mean set and mean boundary are shown

in Figure 6. Here, the expected ODF has no local maxima, minima, or flat regions along the boundary, and therefore both (4.7) and (4.8) are satisfied.

The next two examples show a different way in which the consistency conditions may fail. Example 10 violates (4.7) when $a = 2r$ and the two balls are equally likely. Example 11 violates (4.8).

EXAMPLE 10 (flashing discs). Suppose that $\boldsymbol{A}$ is a disc in $\mathbb{R}^2$ with radius r centred at $(0, 0)$ with probability p, and otherwise it is a disc with radius r centred at $(a, 0)$.

For $p = 0.5$, $E[\boldsymbol{A}]$ is an ellipse if the two discs intersect, and it is empty otherwise. The critical case occurs for $a = 2r$, when $E[\boldsymbol{A}] = E[\partial\boldsymbol{A}] = \{(t, 0) : 0 \leq t \leq a\}$. Here, $E[b_{\boldsymbol{A}}(x)]$ has a local minimum at $E[\partial\boldsymbol{A}]$ and condition (4.7) fails. Figure 7 shows the three possible cases. For $p \neq 0.5$, condition (4.7) is satisfied for $a = 2r$. Figure 7 shows the case when $p < 0.5$.

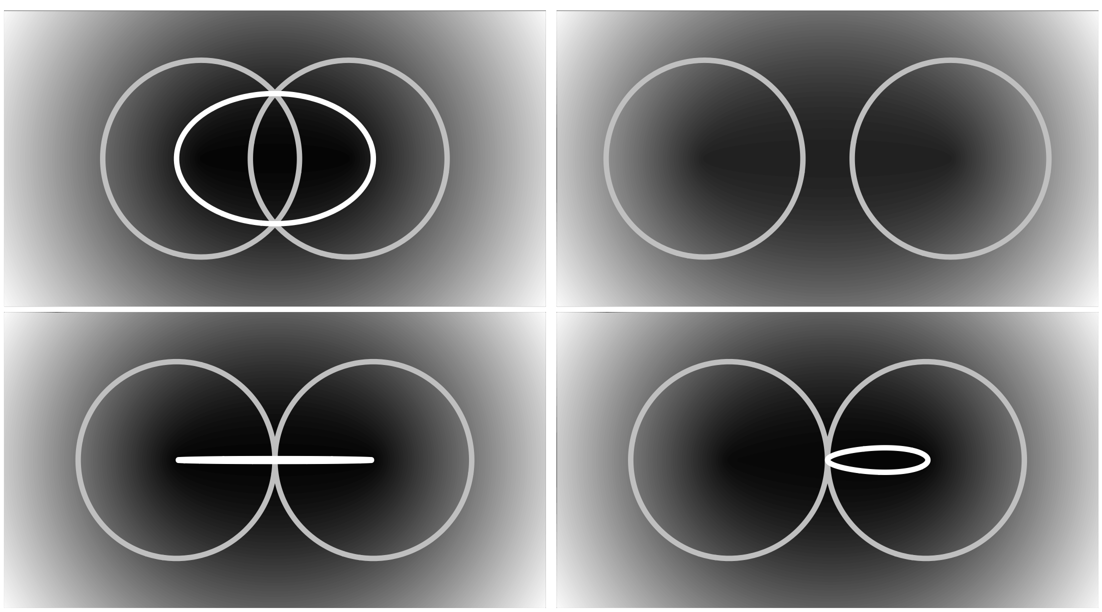

FIGURE 7. The expected boundary of the mean set (white) in Example 10 superimposed on the grey scale image of the expected ODF: (top row) $p = 0.5$ and $a < 2r$ (left), $a > 2r$ (right); (bottom row) $a = 2r$ and $p = 0.5$ (left), $p = 0.49$ (right). The boundaries of the original discs are shown in grey.

EXAMPLE 11 (flashing discs in reverse). For $\mathcal{D} = [-4, 4]^2 \subset \mathbb{R}^2$, let $A_1 = \overline{\mathcal{D} \setminus B_1((0, 0))}$ and $A_2 = \overline{\mathcal{D} \setminus B_1((2, 0))}$. Suppose that we observe either A_1 or A_2

with equal probability. These sets are the complements of the flashing discs from Example 10 at the critical case $a = 2r$. Here, we obtain

$$E[b_{\boldsymbol{A}}(x_1, 0)] \quad = \quad \begin{cases} x_1, & x_1 \leq 0 \\ 0, & x_1 \in [0, 2] \\ -x_1, & x_1 \geq 2, \end{cases}$$

and $E[b_{\boldsymbol{A}}(x_1, x_2)] < 0$ for $x_2 \neq 0$. Hence $E[\boldsymbol{A}] = \mathcal{D}$ and $E[\partial \boldsymbol{A}] = [0, 2] \times [0, 1]$. Therefore, $E[b_{\boldsymbol{A}}(x)]$ has a local maximum along $E[\partial \boldsymbol{A}]$ and does not satisfy condition (4.8).

The preceding examples show that the consistency conditions of Corollary 4.2 can be violated in a number of ways. The following result provides further insight into these conditions.

Proposition 4.5. *Condition* (4.7) *holds iff* $E[\boldsymbol{A}] = \overline{(E[\boldsymbol{A}^c])^c}$. *Condition* (4.8) *holds iff* $E[\boldsymbol{A}^c] = \overline{(E[\boldsymbol{A}])^c}$. *Furthermore, condition* (4.7) *holds iff* $\partial E[\boldsymbol{A}^c] = E[\partial \boldsymbol{A}]$, *and condition* (4.8) *holds iff* $\partial E[\boldsymbol{A}] = E[\partial \boldsymbol{A}]$.

We need to address certain technical details in the above result. As stated, the definition of $E[\boldsymbol{A}]$ is valid for a r.c.s. with non-zero boundary. However, if $\boldsymbol{A}$ is closed, then $\boldsymbol{A}^c$ is not necessarily closed, and the closure of $\boldsymbol{A}^c$ may have an empty boundary (e.g. $\overline{\{0, 1\}^c}$), whereby $b_{\overline{\boldsymbol{A}^c}}$ is not well defined. However, since $b_{\boldsymbol{A}}$ is a well-defined random variable, and since $b_{\boldsymbol{A}^c}(x) = -b_{\boldsymbol{A}}(x)$, then $\boldsymbol{A}^c$ has a well-defined ODF. We therefore use the definition $E[\boldsymbol{A}^c] = \{x : E[b_{\boldsymbol{A}^c}(x)] \leq 0\} = \{x : E[b_{\boldsymbol{A}}(x)] \geq 0\}$.

4.3. **Confidence Sets.** Using the method of Section 3, we obtain confidence regions for the expected set and the expected boundary. In this section, we assume that the observed sets $\boldsymbol{A}_1, \ldots, \boldsymbol{A}_n$ are IID.

Theorem 4.6. *Suppose that* $E[b^2_{\boldsymbol{A}}(x_0)] < \infty$ *for some* $x_0 \in \mathcal{D}$. *Then*

$$\mathbb{Z}_n(x) \equiv \sqrt{n}(\bar{b}_n(x) - E[b_{\boldsymbol{A}}(x)]) \Rightarrow \mathbb{Z}(x),$$

where $\mathbb{Z}$ *is a Gaussian random field with mean zero and covariance*

$$cov(\mathbb{Z}(x), \mathbb{Z}(y)) = E[b_{\boldsymbol{A}}(x) b_{\boldsymbol{A}}(y)] - E[b_{\boldsymbol{A}}(x)] E[b_{\boldsymbol{A}}(y)] \quad \text{for } x, y \in \mathcal{D}.$$

The next result shows that $\mathbb{Z}$ has smooth sample paths. Recall that a function $f : \mathbb{R}^d \mapsto \mathbb{R}$ is Lipschitz of order α if it satisfies

$$|f(x) - f(y)| \leq K|x - y|^\alpha,$$

for some positive finite constant K and $\alpha > 0$, for all x, y in the domain of f.

Proposition 4.7. *For any* $x, y, x', y' \in \mathcal{D}$

$$var(\mathbb{Z}(x) - \mathbb{Z}(y)) \quad \leq \quad |x - y|^2, \tag{4.9}$$

$$|cov(\mathbb{Z}(y) - \mathbb{Z}(x), \mathbb{Z}(y') - \mathbb{Z}(x'))| \quad \leq \quad 2|y - x||y' - x'|. \tag{4.10}$$

Moreover, the sample paths of $\mathbb{Z}$ *are Lipshitz of order* α, *for any* $\alpha < 1$.

We now construct confidence sets for both $E[\boldsymbol{A}]$ and $E[\partial\boldsymbol{A}]$. Let $\mathcal{W} \subset \mathcal{D}$ be a compact set, and define q_1 to be a number such that $\mathrm{pr}(\sup_{x\in\mathcal{W}} \mathbb{Z}(x) \leq q_1) = 1-\alpha$, then

$$\left\{x \in \mathcal{W} : \bar{b}_n(x) \leq \frac{1}{\sqrt{n}}\, q_1\right\} \tag{4.11}$$

is a $100(1-\alpha)\%$ confidence region for $E[\boldsymbol{A}] \cap \mathcal{W} = \{x \in \mathcal{W} : E[b_{\boldsymbol{A}}(x)] \leq 0\}$. Also, let q_2 denote a number such that $\mathrm{pr}(\sup_{x\in\mathcal{W}} |\mathbb{Z}(x)| \leq q_2) = 1-\alpha$, then

$$\left\{x \in \mathcal{W} : |\bar{b}_n(x)| \leq \frac{1}{\sqrt{n}}\, q_2\right\} \tag{4.12}$$

is a $100(1-\alpha)\%$ confidence region for $E[\partial\boldsymbol{A}] \cap \mathcal{W} = \{x \in \mathcal{W} : E[b_{\boldsymbol{A}}(x)] = 0\}$. For $\mathcal{W} = \mathcal{D}$, (4.11) and (4.12) give confidence sets for $E[\boldsymbol{A}]$ or $E[\partial\boldsymbol{A}]$, respectively. The quantiles may be approximated using a bootstrap approach.

To properly define the quantiles q_1 and q_2, we need the process $\mathbb{Z}$ to be continuous with probability one. Proposition 4.7 asserts this, and more: $\mathbb{Z}$ is Lipschitz of order $\alpha < 1$. In particular, this tells us that the process is smoother than Brownian motion. The latter satisfies $var(\mathbb{Z}(t) - \mathbb{Z}(s)) = |t-s|$, and is Lipshitz of order $\alpha < 1/2$. To obtain differentiability of $\mathbb{Z}$, we would need it to be of Lipschitz order $\alpha = 1$, and hence the above proposition provides no information in this direction. However, differentiability can be examined using the results in [CL]. Understanding the path properties of the process $\mathbb{Z}$, such as smoothness, provides information about the variability of the quantiles q_1 and q_2 and therefore also on the tightness of the confidence sets.

Theorem 4.8 (Theorem on page 185 in [CL])**.** *Suppose that $r(x,y) = cov(\mathbb{Z}(x), \mathbb{Z}(y))$ has a continuous mixed derivative $r_{11}(x,y) = \partial_x\partial_y r(x,y)$ satisfying*

$$\triangle_{\delta\delta} r_{11}(x,x) \equiv r_{11}(x+\delta, x+\delta) - 2r_{11}(x, x+\delta) + r_{11}(x,x) \leq \frac{C}{|\log|\delta||^a},$$

for some constants $C > 0, a > 3$ and sufficiently small δ, for all $a \leq x \leq b$ for some $a, b \in \mathbb{R}$. Then $\mathbb{Z}(x)$ has a derivative $\mathbb{Z}'(x)$ which is continuous on $[a,b]$.

Example 12 (continuation of Example 5). Let Θ be a Uniform$[-1,1]$ random variable, and consider $\boldsymbol{A} = \{x \in \mathbb{R} : |x - \Theta| \leq 1\}$. Then, after some calculations, we obtain that

$$r_{11}(x,y) \quad = \quad \begin{cases} 1 + x - y - xy & x, y \in [-1,1] \text{ and } x \leq y, \\ 1 + y - x - xy & x, y \in [-1,1] \text{ and } y \leq x, \\ 0 & \text{otherwise.} \end{cases}$$

Hence, for any $\varepsilon > 0$, there exists a δ, sufficiently small, such that

$$\triangle_{\delta\delta} r_{11}(x,x) \quad = \quad \begin{cases} 2|\delta| - \delta^2 & x \in (-1+\varepsilon, 1-\varepsilon) \\ 0 & x < -1-\varepsilon, \text{ or } x > 1+\varepsilon. \end{cases}$$

Therefore, by Theorem 4.8, $\mathbb{Z}(x)$ is continuously differentiable on compact intervals inside $(-\infty,-1),(-1,1)$ or $(1,\infty)$.

Example 13 (disc in $\mathbb{R}^2$ with random centre). The random set is a disc with radius one centred at $(\Theta,0)$ where Θ ~Uniform$[0,2]$, and suppose that we observe 100 IID random sets from this model. The expected set $E[\boldsymbol{A}]$ is shown in Figure 8. Moreover, since $E[b_{\boldsymbol{A}}(x)] \geq |x-x_0|-1$ where $x_0=(E[\Theta],0)$, it follows that $E[\boldsymbol{A}]$ is contained inside the disc of radius one centered at $(1,0)$. Confidence regions were formed for both $E[\partial\boldsymbol{A}]$ and $E[\boldsymbol{A}]$ using re-sampling techniques to estimate the quantiles of $\sup_{x\in\mathcal{W}}\mathbb{Z}(x)$ and $\sup_{x\in\mathcal{W}}|\mathbb{Z}(x)|$ where the window $\mathcal{W}=[-2,2]\times[-1,3]$. These are illustrated in Figure 8.

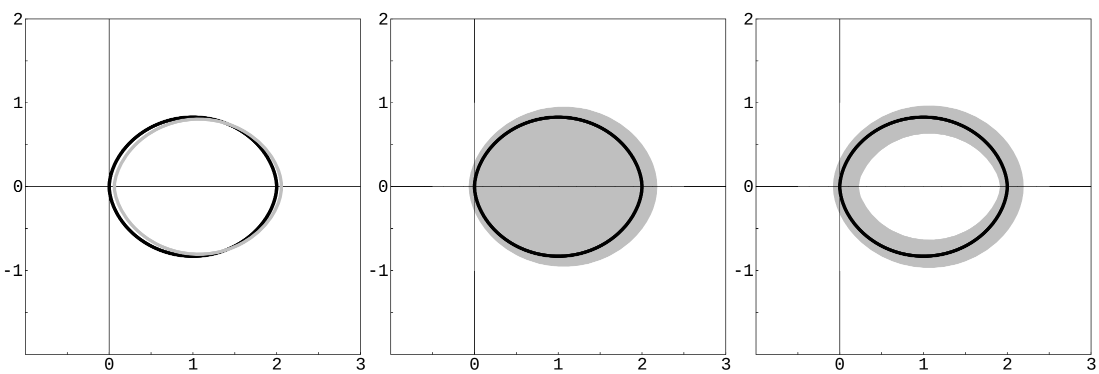

Figure 8. Confidence regions for Example 13: (left) the mean boundary $E[\partial\boldsymbol{A}]$ in black and the empirical boundary $\partial\bar{\boldsymbol{A}}_n$ in grey; (centre) a 95% bootstrap confidence set for $E[\boldsymbol{A}]$; (right) a 95% bootstrap confidence set for $E[\partial\boldsymbol{A}]$. The expected boundary $E[\partial\boldsymbol{A}]$ is shown in black for comparison.

Recall that the confidence set is immune to the consistency conditions (2.3) and (2.4) or (4.7) and (4.8). That is, the empirical set may not be consistent for the mean set $E[\boldsymbol{A}]$, but the confidence set still captures all of $E[\boldsymbol{A}]$ at least $100(1-\alpha)\%$ of the time. We illustrate this point with the following example.

Example 14 (continuation of Example 7). Let $[0,1]\subset\mathcal{D}\subset\mathbb{R}$ and suppose that $\boldsymbol{A}$ is either $\{0,1\}$ or $[0,1]$ with equal probability. Suppose also that we observe a simple random sample of size n from this model. Recall that $E[\boldsymbol{A}]=E[\partial\boldsymbol{A}]=[0,1]$, and $E[b_{\boldsymbol{A}}(x)]$ satisfies neither (4.7) nor (4.8). As before, let $\hat{p}_n$ denote the proportion of times that the set $[0,1]$ is observed. If $\hat{p}_n>0.5$, then $\bar{\boldsymbol{A}}_n=\partial\bar{\boldsymbol{A}}_n=\{0,1\}\neq[0,1]$. If $\hat{p}_n<0.5$, then $\bar{\boldsymbol{A}}_n=[0,1]$ with $\partial\bar{\boldsymbol{A}}_n=\{0,1\}\neq E[\partial\boldsymbol{A}]$.

The fluctuation field is given by

$$\begin{aligned}\mathbb{Z}_n(x) &= \sqrt{n}(\hat{p}_n - 0.5)\left(b_{[0,1]}(x) - b_{\{0,1\}}(x)\right)\\ &\Rightarrow Z\left(b_{[0,1]}(x) - b_{\{0,1\}}(x)\right),\end{aligned}$$

where Z is a univariate normal random variable with mean zero and variance 0.25. The largest difference for $b_{[0,1]}(x) - b_{\{0,1\}}(x)$ occurs at $x = 0.5$, and hence, for any window such that $[0,1] \subset \mathcal{W}$, we have

$$\sup_{x\in\mathcal{W}} \mathbb{Z}(x) = \max\{-Z, 0\}, \quad \text{and} \quad \sup_{x\in\mathcal{W}} |\mathbb{Z}(x)| = |Z|.$$

Therefore, the exact quantiles are $q_1 = 0.5{\cdot}1.645$ and $q_2 = 0.5{\cdot}1.96$, and the confidence region for $E[\partial \boldsymbol{A}]$ is given by $\{x : |\bar{b}_n(x)| \le 0.5 \cdot 1.96/\sqrt{n}\}$.

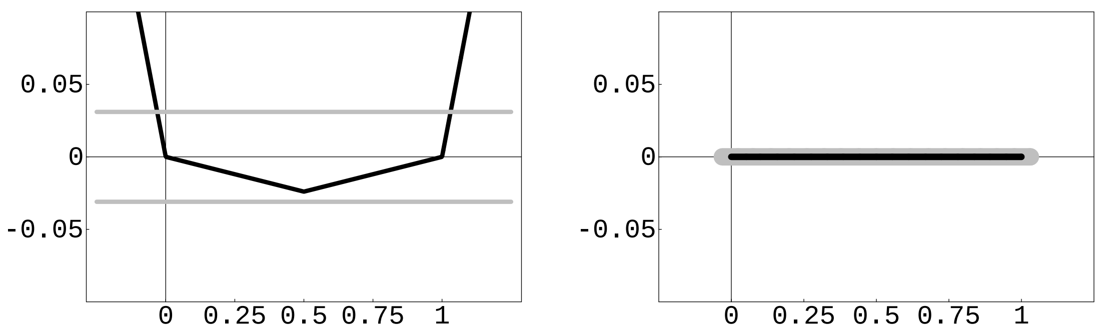

FIGURE 9. Confidence sets for Example 7: the empirical ODF $\bar{b}_n$ with the quantile levels $-q_2, q_2$ (left) and the confidence set in grey with $E[\partial \boldsymbol{A}]$ shown in black (right).

Figure 9 illustrates the formation of the confidence set in this case for $n = 1000$. In this example, the proportion of times that $[0,1]$ was observed is $\hat{p}_n = 0.489$ and hence $\partial\bar{\boldsymbol{A}}_n = \{0,1\}$. The left panel of Figure 9 shows the observed function $\bar{b}_n(x)$ (black) along with the lower and upper quantile levels $\{-q_2, q_2\} = \{-0.5 \cdot 1.96/\sqrt{1000}, 0.5 \cdot 1.96/\sqrt{1000}\}$ (grey). The confidence interval for $E[\partial \boldsymbol{A}] = [0,1]$ is then $(-0.031, 1.031)$ (Figure 9, right).

Now, for any n, $\max_{x\in[0,1]} |\bar{b}_n(x)| = |\bar{b}_n(0.5)| = |0.5 - \hat{p}_n|$. Therefore the confidence region misses a part of $E[\partial \boldsymbol{A}] = [0,1]$ if and only if $|0.5 - \hat{p}_n| > 0.5 \cdot 1.96/\sqrt{n}$. This happens with probability 0.95, for sufficiently large n. On the other hand, the Hausdorff distance $\rho(E[\partial \boldsymbol{A}], \partial\bar{\boldsymbol{A}}_n) = 0.5$ whenever $\hat{p}_n \neq 0.5$.

4.4. **Separable Random Closed Sets.** As in [SJ], we say that a random closed set is separable if there exists a random variable Θ and functions h_j, g_j, $j = 1, \ldots, k$ such

that

$$b_{\boldsymbol{A}}(x) \;\equiv\; \sum_{j=1}^{k} h_j(x) g_j(\Theta) \;\text{ almost surely.}$$

For example, the ball centered at x_0 with random radius R is separable, as $b_{\boldsymbol{A}}(x) = |x - x_0| - R$. Notably, the mean of a separable random set has the same geometric structure as the original sets. The same is true of the expected boundary. In this section, we investigate the confidence regions for this special class of random closed sets.

To this end, suppose that we observe IID samples of the random variable Θ, and let $\bar{g}_{n,j} = 1/n \sum_{i=1}^{n} g_j(\Theta_i)$. We assume that $E[g_j(\Theta)^2] < \infty$ for all $j = 1, \ldots, k$. It follows immediately that $E[b_{\boldsymbol{A}}(x)] = \sum_{j=1}^{k} h_j(x) E[g_j(\Theta)]$ and that

$$\mathbb{Z}_n(x) = \sum_{j=1}^{k} \sqrt{n}(\bar{g}_{n,j} - E[g_j(\Theta)])\, h_j(x) \Rightarrow \sum_{j=1}^{k} Z_j h_j(x),$$

where $Z = \{Z_1, \ldots, Z_k\}$ is a multivariate normal random variable with mean zero and variance matrix given by $\operatorname{cov}(Z_j, Z_m) = \operatorname{cov}(g_j(\Theta), g_m(\Theta))$.

Remark 4.9. *Consider the special case $b_{\boldsymbol{A}}(x) = h(x) + g(\theta)$. Then $\bar{\boldsymbol{A}}_n = \{x : h(x) + \bar{g}_n \leq 0\}$, $\partial\bar{\boldsymbol{A}}_n = \{x : h(x) + \bar{g}_n = 0\}$, $E[\boldsymbol{A}] = \{x : h(x) + E[g(\Theta)] \leq 0\}$ and $E[\partial\boldsymbol{A}] = \{x : h(x) + E[g(\Theta)] \leq 0\}$. Note that by definition $h(x)$ is continuous. If $h(x)$ satisfies condition* (2.3) *at $p = -E[g(\Theta)]$, then $\bar{\boldsymbol{A}}_n$ converges strongly to $E[\boldsymbol{A}]$. In addition, if $h(x)$ satisfies condition* (2.4) *at $p = -E[g(\Theta)]$, then $\partial\bar{\boldsymbol{A}}_n$ converges strongly to $E[\partial\boldsymbol{A}]$. Also, $\mathbb{Z}_n(x) = \sqrt{n}(\bar{g}_n - E[g(\Theta)])$, which converges to $\mathbb{Z} \sim Normal(0, var(g(\Theta)))$. Thus, q_1 and q_2 are easily calculated from the quantiles of the univariate normal distribution.*

Example 15 (confidence set for disc with random radius). Suppose that $\boldsymbol{A}$ is a disc with random radius R with $\mu = E[R]$ and $\sigma^2 = \operatorname{var}(R)$. Then the expected set is a circle with radius μ. Also, the 95% confidence interval for $E[\boldsymbol{A}]$ is a circle with radius $\mu + 1.645\sigma/\sqrt{n}$, while the 95% confidence set for $E[\partial\boldsymbol{A}]$ is the band $\{x : \mu - 1.96\sigma/\sqrt{n} \leq |x| \leq \mu + 1.96\sigma/\sqrt{n}\}$.

Example 16 (upper half-plane at random angle). Suppose that $\boldsymbol{A} = \{x : x_2 \geq x_1 \tan\Theta\}$, the upper half plane making angle Θ with the x_1-axis, and suppose that $\Theta \sim$Uniform[a,b] with $0 < b - a < 2\pi$. Then $b_{\boldsymbol{A}}(x) = x_1 \sin(\Theta) - x_2 \cos(\Theta)$ and

$$E[b_{\boldsymbol{A}}(x)] \;=\; \frac{2}{b-a} \sin\left(\frac{b-a}{2}\right) \left\{ x_1 \sin\left(\frac{b+a}{2}\right) - x_2 \cos\left(\frac{b+a}{2}\right) \right\}$$

so that $E[\boldsymbol{A}] = \left\{x : x_2 \geq x_1 \tan\left(\frac{a+b}{2}\right)\right\}$. Because at least one of $\sin((a+b)/2)$ or $\cos((a+b)/2)$ is non-zero, the expected ODF $E[b_{\boldsymbol{A}}(x)]$ is linear (and non-constant), and therefore satisfies both of the consistency conditions.

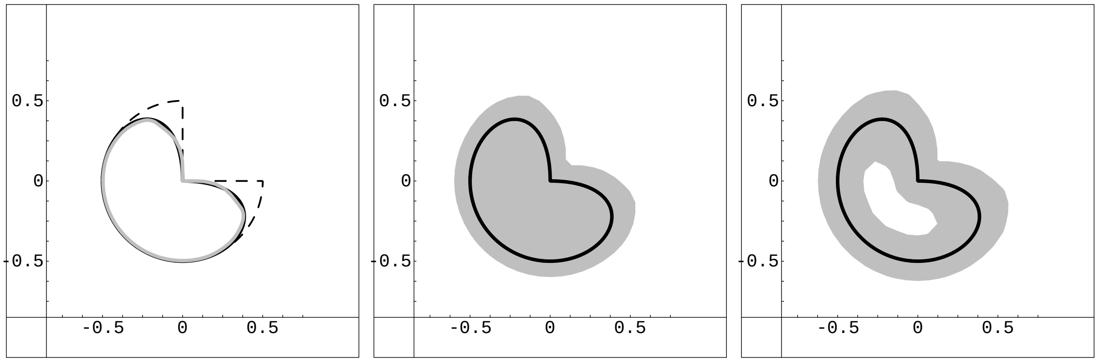

FIGURE 10. Left: the expected boundary $E[\partial \boldsymbol{A}]$ (black), its estimate based on 25 samples (grey) and the boundary of a pacman with radius 0.5 (dashed); centre and right: 95% bootstrapped confidence regions for $E[\boldsymbol{A}]$ and $E[\partial \boldsymbol{A}]$, respectively. The confidence sets are denoted by the shaded area, while the black line shows $E[\partial \boldsymbol{A}]$.

Next, the limiting fluctuation process is $\mathbb{Z}(x) = x_1 Z_1 - x_2 Z_2$, where $\mathrm{cov}(Z_1, Z_2) = \mathrm{cov}(\cos(\Theta), \sin(\Theta))$. For a fixed window $\mathcal{W}$, the maximum of $\mathbb{Z}$ occurs on the boundary with probability one. Therefore, the variability of $\mathbb{Z}$ depends on the radius of $\mathcal{W}$, $\max_{x\in\mathcal{W}} |x|$. Suppose that $\mathcal{W}_1 = \{x : |x_1| \leq 1, |x_2| \leq 1\}$ and $\mathcal{W}_2 = \{x : |x_1| \leq 2, |x_2| \leq 2\}$. We then have

$$\max_{x\in\mathcal{W}_1} \mathbb{Z}(x) = \max\{Z_1 - Z_2, Z_1 + Z_2, -Z_1 - Z_2, -Z_1 + Z_2\} = |Z_1| + |Z_2|,$$

and $\max_{x\in\mathcal{W}_2} \mathbb{Z}(x) = 2 \max_{x\in\mathcal{W}_1} \mathbb{Z}(x)$. For the special case of $a = 0, b = \pi$, we obtain that $E[b_{\boldsymbol{A}}(x)] = 2x_1/\pi, E[\boldsymbol{A}] = \{x : x_1 \geq 0\}$ and the variance of $Z = \{Z_1, Z_2\}$ is equal to $I/2$, where I is the 2×2 identity matrix. From simulations, we find $P(\max_{x\in\mathcal{W}_1} \mathbb{Z}(x) \leq 2.23) = 0.95$. Therefore, 95% confidence regions for the set $E[\boldsymbol{A}]$ over the windows $\mathcal{W}_1$ and $\mathcal{W}_2$ are

$$\{x \in \mathcal{W}_1 : \bar{b}_n(x) \leq 2.23/\sqrt{n}\}, \quad \text{and} \quad \{x \in \mathcal{W}_2 : \bar{b}_n(x) \leq 4.46/\sqrt{n}\}.$$

respectively. As expected, the smaller window $\mathcal{W}_1$ yields a tighter confidence region.

EXAMPLE 17 (pacman in $\mathbb{R}^2$). Define the pacman with radius r, $A(r)$, to be a disc with radius r centred at the origin with its upper left quadrant removed. That is,

$$A(r) = \{x : |x| \leq r\} \cap \{\{x : x_1 \leq 0\} \cup \{x : x_2 \leq 0\}\}.$$

Figure 4 (right) shows the ODF of $A(r)$ for $r = 1$.

Suppose that $\boldsymbol{A} = \boldsymbol{A}(R)$, where R is a uniform random variable on $[0, 1]$. Then the expected set $E[\boldsymbol{A}]$ is a smoothed version of $A(0.5)$, as seen in Figure 10 (left). The figure also shows bootstrapped 95% confidence sets for both $E[\boldsymbol{A}]$ and $E[\partial \boldsymbol{A}]$.

The accuracy of the estimate and the apparent centering of the confidence intervals around the mean set may be explained upon closer inspection of the sample. The ODF of the pacman is similar to that of the circle; indeed, they are identical in the lower left quadrant. Therefore, the behaviour of the estimators and of the confidence regions is not unlike that of estimators and confidence intervals for the real-valued $E[R] = 0.5$. For our sample of $n = 25$, we observed $\bar{R} = 0.496$, which explains the accuracy of the estimator $\partial \bar{A}_n$. On the other hand, the confidence region still shows the large variability of $\partial \bar{A}_n$.

5. Further Examples

5.1. **Application to Image Reconstruction.** Image averaging arises in various situations, for example, when multiple images of the same scene are observed or when the acquired images represent objects of the same class and the goal is to determine the average object (shape) that can be described as typical. Here we consider the example of image averaging studied in [BM]. An original newspaper image (see Figure 11) was contaminated and then reconstructed using an Ising prior model. The details are given in [BM], Section 6. The data consists of 15 independent binary images of the reconstructed image.

Figure 11. From left to right are shown the original image, the distance-average reconstruction and the ODF-average reconstruction.

In [BM], the original image (Figure 11, left) was reconstructed using the distance-average approach based on the ODF; see Figure (centre) and Figure 6 in [BM]. The ODF-reconstruction using the method proposed in [SJ] is also shown in Figure 11 (right). This reconstruction, which we call the ODF-average here, outperforms the distance-average both in terms of misclassification error and the L_2-distance consistent with the distance-average construction [SJ].

Next, we compute 95% confidence regions for the ODF-average reconstruction based on 5K bootstrap samples. Figure 12 (left top – full image, left bottom – inset) shows the confidence set for $E[\boldsymbol{A}]$ with the boundary of the true image overlayed in black. The confidence set contains all of the true text image and appears tight, although there are a number of spurious bounds induced by noise. The confidence set

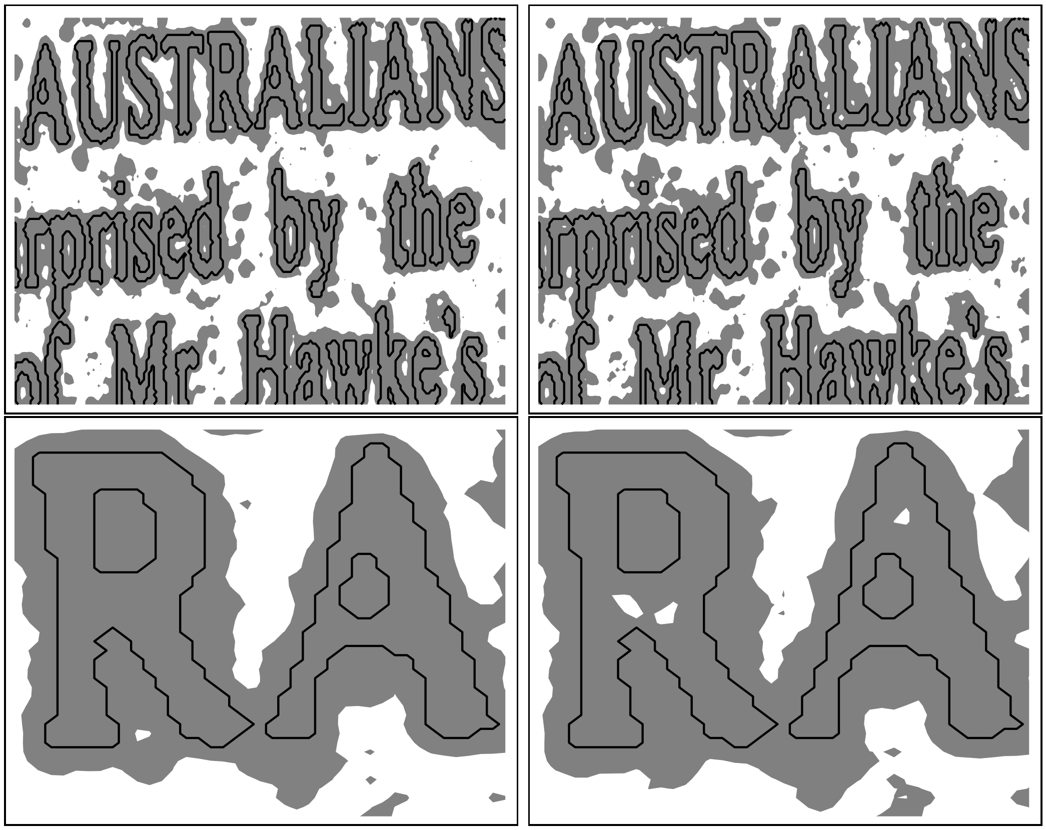

Figure 12. Confidence regions for the expected set (left) and the expected boundary (right) shown with the boundary of the true image (black). The bottom row shows insets of the original image.

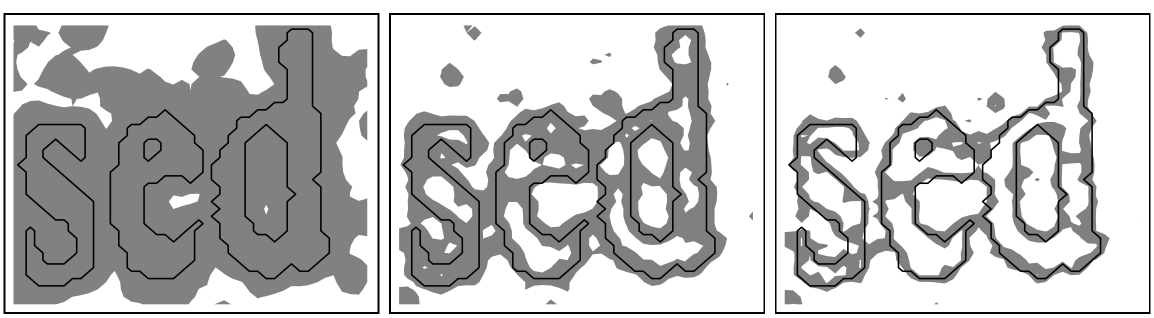

Figure 13. Confidence regions for the expected boundary with the boundary of the true image (green) using the true sample size (left) and hypothetical sample sizes of $n = 50, 100$ (middle and right, respectively). The pictures shown are insets of the full image.

for $E[\partial \boldsymbol{A}]$ is also shown in Figure 12 (right top – full image, right bottom – inset). The boundary of the lower confidence set consists of only a few closed contours scattered throughout the text. The lack of tightness in the confidence set is explained by the small sample size and a relatively thin font width. Hypothetically increasing the sample size would produce tighter confidence sets for both the expected set and its boundary. For example, the bootstrap confidence set for the boundary based on 50 and 100 samples is tighter as compared to the one based on 15 samples (see Figure 13). It should be noted that confidence intervals in Figure 13 were created using the original window $\mathcal{D}$, and that the picture is a close-up of the result.

5.2. **Application to Medical Imaging.** We next consider an example of boundary reconstruction in mammography, where the skin-air contour is used to determine the radiographic density of the tissue and to estimate breast asymmetry. Both measures are known to be associated with the risk of developing breast cancer [SMW$^+$, DWW$^+$].

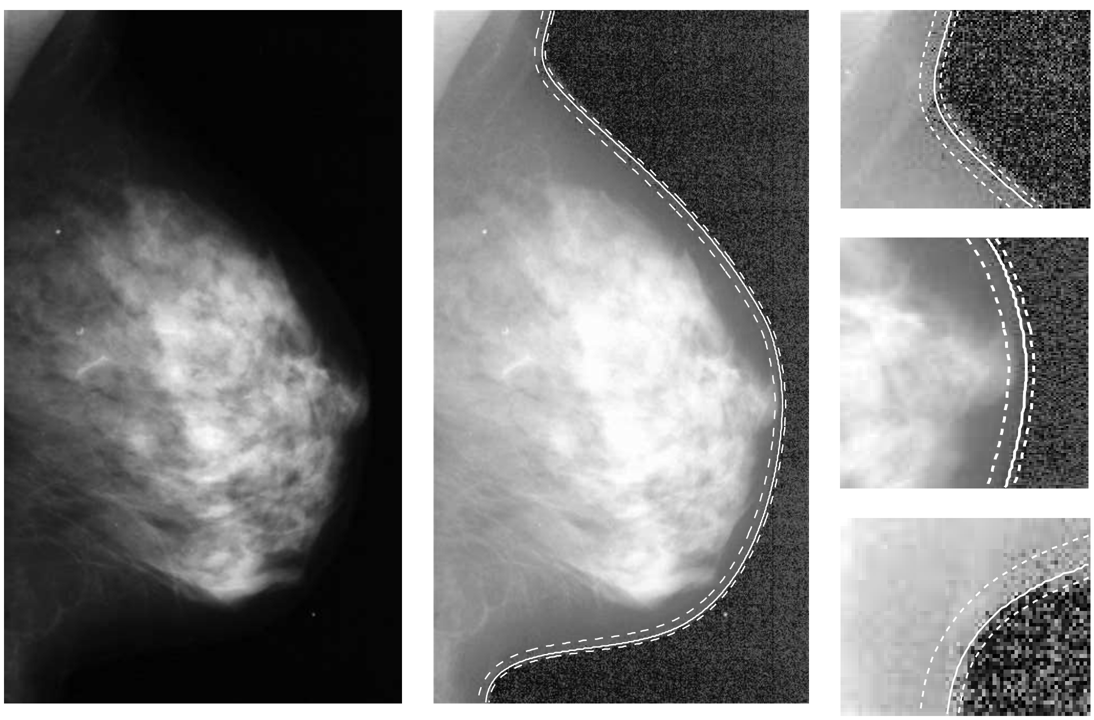

Figure 14. Confidence sets for the reconstructed skin-air boundary in a mammogram: the original image (left), and the digitally enhanced image (centre) with the reconstructed boundary (solid line) and confidence region (dashed line). Three insets are also shown (right).

In [SB], B-spline curves were used to reconstruct a smooth connected boundary of an object in a noisy image, and the method was applied to estimate the tissue boundary in mammograms. The method is Bayesian and uses a loss function based on ODFs to determine the optimal estimator for the boundary. More details on the reconstruction method and image acquisition can be found in [SB]. Here, we apply the proposed method to construct a confidence set for the boundary estimator, which is given by the zero-level isocontour of the curve samples from the posterior. We emphasize that the confidence region is not a credible set, but rather it describes the variability of the sample curves.

Figure 14 (left) shows a typical digitized mammogram image, characterised by a low contrast-to-noise ratio. A probability integral transform improves the contrast by increasing the dynamic range of image intensities (Figure 14, centre). The 95% confidence set (dashed) for the boundary estimator (solid) in Figure 14 is obtained using a bootstrap resampling of size 1000. The confidence set is tight and fits the image well. It also shows that the reconstructed boundary is more variable toward the inside of the breast tissue. Note that what appears to be a nipple is, in fact, a duct system leading to the nipple, so that the estimators correctly follow the skin line. More details can be seen in insets in Figure 14 (right).

Note that the method proposed in this paper assumes that the observed sets are independent and identically distributed, whilst the boundary reconstruction is based on Monte Carlo sampling from the posterior. To ensure the independence of the curve samples, we construct the confidence set for the boundary using 100 samples from the posterior, which were acquired every 250th sweep after a burn-in period of 1000 sweeps. It remains open to extend the method of confidence sets to dependent samples, in particular, in the context of Bayesian inference.

Appendix

This section contains certain technical details, as well as proofs of the results discussed in this paper.

6.3. **Independent Random Closed Sets.**

In Remark 4.3 we define two r.c.s. $\boldsymbol{A}$ and $\boldsymbol{B}$ to be independent if their ODFs are independent as random functions on $\mathcal{D}$. In this section we give a brief discussion of this definition.

Recall that in [Mat] (page 40), two sets $\boldsymbol{A}$ and $\boldsymbol{B}$ are said to be independent if

$$P(\boldsymbol{A} \cap K_1 \neq \emptyset \text{ and } \boldsymbol{B} \cap K_2 \neq \emptyset) = P(\boldsymbol{A} \cap K_1 \neq \emptyset)P(\boldsymbol{B} \cap K_2 \neq \emptyset),$$

for any compact sets $K_1, K_2 \subset \mathcal{D}$. To differentiate it from our definition, we call this M-independence.

Proposition 6.1. *The relationship between independence and M-independence is as follows.*

1. *Two sets* $\boldsymbol{A}$ *and* $\boldsymbol{B}$ *are M-independent if and only if their distance functions,* $d_{\boldsymbol{A}}(x)$ *and* $d_{\boldsymbol{B}}(x)$*, are independent random functions.*
2. *Two sets* $\boldsymbol{A}$ *and* $\boldsymbol{B}$ *are independent if and only if the sets* $\boldsymbol{A}$ *and* $\boldsymbol{B}$*, and* $\partial\boldsymbol{A}$ *and* $\partial\boldsymbol{B}$*, are both M-independent.*
3. *Two boundary sets* $\partial\boldsymbol{A}$ *and* $\partial\boldsymbol{B}$ *are M-independent if and only if* $|b_{\boldsymbol{A}}(x)|$ *and* $|b_{\boldsymbol{B}}(x)|$ *are independent random functions.*

Thus independence is a stronger notion than M-independence. Recall that $d_A(x) \equiv d_B(x)$ iff $\overline{A} = \overline{B}$, whereas $b_A(x) \equiv b_B(x)$ iff $\overline{A} = \overline{B}$ *and* $\partial A = \partial B$ (e.g. consider the sets $A = [0,1] \cap \mathbb{Q} \subset \mathbb{R}$ and $B = [0,1]$; then $\overline{A} = \overline{B}$, but $\partial A \neq \partial B$, and hence $d_A \equiv d_B$ but $b_A \neq b_B$). Thus, the ODF encapsulates more information about a set than the distance function, and hence more information is required to ascertain its independence.

Proof. We first prove the first part of the statement. Suppose that $d_{\boldsymbol{A}}$ and $d_{\boldsymbol{B}}$ are independent random functions. Then, since for any compact K,

$$\{\boldsymbol{A} \cap K \neq \emptyset\} = \{\inf_{x\in K} d_{\boldsymbol{A}}(x) \leq 0\},$$

it follows that $\boldsymbol{A}$ and $\boldsymbol{B}$ are also M-independent. For the other direction, suppose that $\boldsymbol{A}$ and $\boldsymbol{B}$ are M-independent. Then by the relation

$$\{d_{\boldsymbol{A}}(x) \leq \alpha\} = \{\boldsymbol{A} \cap B_\alpha(x) \neq \emptyset\},$$

for any $\alpha \in \mathbb{R}$, the random functions $d_{\boldsymbol{A}}$ and $d_{\boldsymbol{B}}$ are also independent.

The second part of the proposition is proved in a similar manner. Here, the key relations are

$$\begin{aligned} \{\boldsymbol{A} \cap K \neq \emptyset\} &= \{\inf_{x\in K} b_{\boldsymbol{A}}(x) \leq 0\}, \\ \{\partial\boldsymbol{A} \cap K \neq \emptyset\} &= \{\inf_{x\in K} |b_{\boldsymbol{A}}(x)| = 0\}, \end{aligned}$$

as well as

$$\begin{array}{lll} \{b_{\boldsymbol{A}}(x) \leq \alpha\} = & \{\boldsymbol{A} \cap B_\alpha(x) \neq \emptyset\} & \text{if } \alpha \geq 0, \\ \{b_{\boldsymbol{A}}(x) < \alpha\} = & \{\partial\boldsymbol{A} \cap B_{|\alpha|}(x) \neq \emptyset\}^c \cap \{\boldsymbol{A} \cap \{x\} \neq \emptyset\} & \text{if } \alpha < 0. \end{array}$$

The third statement follows from the first statement, since for any set A, we have $|b_A(x)| = d_{\partial A}(x)$. This completes the proof. □

6.4. **Proofs.**

Proof of Proposition 2.1. Suppose first that (2.4) holds. Then, by the continuity of f,

$$\begin{aligned} \partial\{x : f(x) \leq p\} &= \{x : f(x) \leq p\} \cap \overline{\{x : f(x) > p\}} \\ &= \{x : f(x) = p\} \cap \overline{\{x : f(x) > p\}} \\ &= \{x : f(x) = p\} \cap \{x : f(x) \geq p\} \\ &= \{x : f(x) = p\}. \end{aligned}$$

On the other hand, if $\partial\{x : f(x) \le p\} = \{x : f(x) = p\}$ then $\{x : f(x) = p\} = \overline{\{x : f(x) > p\}} \cap \{x : f(x) = p\}$. Therefore, appealing again to the continuity of f, we find that

$$\begin{aligned}
&\overline{\{x : f(x) < p\}} \\
&= \left(\overline{\{x : f(x) < p\}} \cap \{x : f(x) < p\}\right) \cup \left(\overline{\{x : f(x) < p\}} \cap \{x : f(x) = p\}\right) \\
&= \{x : f(x) < p\} \cup \{x : f(x) = p\} \\
&= \{x : f(x) \le p\},
\end{aligned}$$

as desired. The proof of the first claim is similar, and we omit the details. □

Lemma 6.2. *Suppose that f is continuous. Then*

$$\{x : p_1 \le f(x)\}^\varepsilon \cap \{x : f(x) \le p_2\}^\varepsilon = \{x : p_1 \le f(x) \le p_2\}^\varepsilon.$$

Proof. Suppose y is in the set

$$\{x : p_1 \le f(x)\}^\varepsilon \cap \{x : f(x) \le p_2\}^\varepsilon \setminus \{x : p_1 \le f(x) \le p_2\}.$$

Then one of two possibilities exists: Either $f(y) < p_1$ or $f(y) > p_2$. The argument for both cases is the same, so we present only the first instance.

Assume then that y is such that $f(y) < p_1$. By definition of y, there exists an x_1 such that $p_1 \le f(x_1)$ and $y \in B_\varepsilon(x_1)$, or an x_2 such that $f(x_2) \le p_2$ and $y \in B_\varepsilon(x_2)$. In the first setting, since f is continuous, there also exists a z such that $p_1 \le f(z) \le p_2$ and $d(y, z) \le \varepsilon$. For example, one such z must fall on the line between y and x_1, which would clearly satisfy $|y - z| \le \varepsilon$. A similar argument shows that if $y \in B_\varepsilon(x_2)$, then there exists a z such that $p_1 \le f(z) \le p_2$ and $|y - z| \le \varepsilon$. It follows that $y \in B(z, \varepsilon)$. This proves that

$$\{x : p_1 \le f(x)\}^\varepsilon \cap \{x : f(x) \le p_2\}^\varepsilon \subset \{x : p_1 \le f(x) \le p_2\}^\varepsilon.$$

Containment in the other direction is immediate, completing the proof. □

Proof of Theorem 2.3. The proof here is similar to that of [Mol]. If $f : \mathcal{D} \mapsto \mathbb{R}$ is a continuous function satisfying the conditions of the theorem, then

$$\begin{aligned}
\widetilde{\varphi}(\pm\varepsilon) &= \rho(\{x : f(x) \le p_2\}, \{x : f(x) \le p_2 \pm \varepsilon\}) \\
\varphi(\pm\varepsilon) &= \rho(\{x : p_1 \le f(x)\}, \{x : p_1 \pm \varepsilon \le f(x)\})
\end{aligned}$$

are all continuous for ε near zero, and moreover, they both converge to zero as $\varepsilon \to 0$. Now, by (A1), we know that $\widehat{f}_n$ converges uniformly to f with probability one. Let

$$\eta_n = \sup_{x \in \mathcal{D}} |f(x) - \widehat{f}_n(x)|,$$

and also define

$$\varepsilon_n = \max\{\varphi(\eta_n), \varphi(-\eta_n), \widetilde{\varphi}(\eta_n), \widetilde{\varphi}(-\eta_n)\}$$

which converges to zero as $n \to \infty$ almost surely. We will next show that $\rho(\{x : p_1 \leq \widehat{f}_n(x) \leq p_2\}, \{x : p_1 \leq f(x) \leq p_2\}) \leq \varepsilon_n$. To this end

$$\begin{aligned}\{x : p_1 \leq f(x) \leq p_2\} &\subset \{x : f(x) \leq p_2 - \eta_n\}^{\widetilde{\varphi}(-\eta_n)} \\ &\subset \{x : \widehat{f}_n(x) \leq p_2\}^{\widetilde{\varphi}(-\eta_n)} \\ &\subset \{x : \widehat{f}_n(x) \leq p_2\}^{\varepsilon_n}.\end{aligned}$$

Repeating in the other direction, we obtain

$$\begin{aligned}\{x : p_1 \leq f(x) \leq p_2\} &\subset \{x : p_1 + \eta_n \leq f(x)\}^{\varphi(\eta_n)} \\ &\subset \{x : p_1 \leq \widehat{f}_n(x)\}^{\varphi(\eta_n)} \\ &\subset \{x : p_1 \leq \widehat{f}_n(x)\}^{\varepsilon_n}.\end{aligned}$$

and hence, by Lemma 6.2,

$$\{x : p_1 \leq f(x) \leq p_2\} \subset \{x : p_1 \leq \widehat{f}_n(x) \leq p_2\}^{\varepsilon_n}. \tag{A-1}$$

For the other direction,

$$\begin{aligned}\{x : p_1 \leq \widehat{f}_n(x) \leq p_2\} &\subset \{x : f(x) \leq p_2 + \eta_n\} \\ &\subset \{x : f(x) \leq p_2\}^{\varepsilon_n}.\end{aligned}$$

A similar argument shows that $\{x : p_1 \leq \widehat{f}_n \leq p_2\} \subset \{x : p_1 \leq f(x)\}^{\varepsilon_n}$, from which it follows

$$\{x : p_1 \leq \widehat{f}_n(x) \leq p_2\} \subset \{x : p_1 \leq f(x) \leq p_2\}^{\varepsilon_n}$$

by Lemma 6.2. Together with (A-1) this proves the result.

To address necessity, suppose that there exists a neighbourhood of x_0, $B_\delta(x_0)$, and a subsequence n_k such that $\widehat{f}_{n_k}(x) < f(x)$ for all $x \in B_\delta(x_0)$. Assume also that $x_0 \in \{x : p_1 \leq f(x)\} \setminus \overline{\{x : p_1 < f(x)\}}$. In particular, this implies that (2.4) is not satisfied, and hence there exists an $\varepsilon > 0$ such that $\rho(x_0, \overline{\{x : p_1 < f(x) \leq p_2\}}) > \varepsilon$. It follows that $\rho(\widehat{F}_{n_k}(p_1, p_2), F(p_1, p_2)) > \min(\varepsilon, \delta) > 0$, proving the result. A similar argument proves the other claim. □

Proof of Proposition 2.5. Without loss of generality, we assume that $p = 0$. For any fixed $x_0 \in \mathcal{D}$, and for $t \in \mathbb{R}_+$ define the function $h(t) = f(x_0 + te_0)$. By assumption, it is a strictly increasing function in t. Let t^* denote a value such that

$$h(t^*) = 0.$$

Then for all $t > t^*$ we have that $x_0 + te_0 \in \{x : f(x) > 0\}$ and for all $t < t^*$ we have that $x_0 + te_0 \in \{x : f(x) < 0\}$. Note also that for all $x \in \partial F(0) = \{x \in \mathcal{D} : f(x) = 0\}$, there exists an x_0 such that $x_0 + te_0 = x$ for some t. It follows that for all $x \in \partial F(0)$ and for all $\varepsilon > 0$, $B_\varepsilon(x) \cap \{x \in \mathcal{D} : f(x) < 0\} \neq \emptyset$ and $B_\varepsilon(x) \cap \{x \in \mathcal{D} : f(x) > 0\} \neq \emptyset$. Therefore,

$$\overline{\{x : f(x) < 0\}} = \{x : f(x) \leq 0\} \quad \text{and} \quad \overline{\{x : f(x) > 0\}} = \{x : f(x) \geq 0\},$$

as required. □

Proof of Remark 2.7. The extension to unbounded domains of the consistency results is immediate, because strong convergence of A_n to A is defined as

$$\rho(A_n \cap K, A \cap K) \to 0, \quad \text{a.s. as} \quad n \to \infty,$$

for each compact set $K \subset \mathbb{R}^d$. □

Proof of Corollary 4.2. Since the functions $\bar{b}_n(x)$ and $E[b_{\boldsymbol{A}}(x)]$ are continuous, the result follows directly from Theorems 2.2 and 2.3. □

Proof of Remark 4.3. It is well known that the oriented distance function is uniformly Lipshitz [DZ2]. That is, for a fixed set A,

$$|b_A(x) - b_A(y)| \le |x - y|, \quad \text{for all } x, y \in \mathcal{D}. \tag{A-2}$$

This property immediately follows also for $E[b_{\boldsymbol{A}}(x)]$ and for the empirical ODF, $\bar{b}_n(x)$. It follows also that if $E[|b_{\boldsymbol{A}}(x_0)|] < \infty$ for some $x_0 \in \mathcal{D}$, then the mean $E[b_{\boldsymbol{A}}(x)]$ exists for all $x \in \mathcal{D}$. Moreover, since the window $\mathcal{D} \subset \mathbb{R}^d$ is compact, we immediately obtain the uniform convergence of $\bar{b}_n(x)$. That is,

$$\lim_{n\to\infty} \sup_{x\in\mathcal{D}} |\bar{b}_n(x) - E[b_{\boldsymbol{A}}(x)]| = 0,$$

as required. □

Proof of Proposition 4.5. From the definition of the oriented distance function, we have that $b_{\boldsymbol{A}}(x) = -b_{\boldsymbol{A}^c}(x)$ almost surely. Therefore $E[b_{\boldsymbol{A}}(x)] = -E[b_{\boldsymbol{A}^c}(x)]$ and hence

$$E[\boldsymbol{A}^c] \;\; = \;\; \{x : E[b_{\boldsymbol{A}}(x)] \ge 0\}.$$

Therefore we have the following relation

$$\overline{(E[\boldsymbol{A}])^c} = \overline{\{x : E[b_{\boldsymbol{A}}(x)] > 0\}} = \{x : E[b_{\boldsymbol{A}}(x)] \ge 0\} = E[\boldsymbol{A}^c],$$

which holds iff (4.8) is satisfied. Similarly,

$$\overline{(E[\boldsymbol{A}^c])^c} = \overline{\{x : E[b_{\boldsymbol{A}}(x)] < 0\}} = \{x : E[b_{\boldsymbol{A}}(x)] \le 0\} = E[\boldsymbol{A}],$$

which holds iff (4.7) is satisfied. The last statement of the proposition is a direct corollary of Proposition 2.1. □

Theorem 6.3 (Theorem 1.4.7 on page 38 in [Kun]: Kolmogorov's tightness criterion.)**.** *For a compact set $\mathcal{D} \subset \mathbb{R}^d$, let $\{Y_n(x) : x \in \mathcal{D}\}$ be a sequence of continuous random fields with values in $\mathbb{R}$. Assume that there exist positive constants γ, C and $\alpha_1, \dots, \alpha_d$ with $\sum_{i=1}^d \alpha_i^{-1} < 1$ such that*

$$\begin{aligned} E[|Y_n(x) - Y_n(y)|^\gamma] \;&\le\; C\left(\sum_{i=1}^d |x_i - y_i|^{\alpha_i}\right) \qquad \text{for every } x, y \in \mathcal{D}, \\ E[|Y_n(x)|^\gamma] \;&\le\; C, \qquad \text{for all } x \in \mathcal{D}, \end{aligned}$$

holds for any n. Then $\{Y_n\}$ is tight in $C(\mathcal{D})$.

Lemma 6.4. *Consider the process $\mathbb{Z}_n(x) = \sqrt{n}(\bar{b}_n(x) - E[b_{\boldsymbol{A}}(x)])$ on $x \in \mathcal{D}$ defined in Section 4.3. Then there exists a constant $C(d)$, depending only on d, such that*

$$E\left[|\mathbb{Z}_n(x) - \mathbb{Z}_n(y)|^{2d}\right] \;\; \leq \;\; C(d)|x-y|^{2d},$$

for any n and $x, y \in \mathcal{D}$.

Proof of Lemma 6.4. The case $d = 1$ is immediate. Next, consider $d = 2$,

$$E\left[|\mathbb{Z}_n(x) - \mathbb{Z}_n(y)|^4\right] \;\; = \;\; n^{-2} \sum_{i,j,k,l=1}^{n} E\left[b_i^* b_j^* b_k^* b_l^*\right],$$

where $b_i^* = b_i(x) - b_i(y) - E[b_{\boldsymbol{A}}(x)] + E[b_{\boldsymbol{A}}(y)]$, and $|b_i^*| \leq 2|x-y|$ almost surely, since both b_i and $E[b_{\boldsymbol{A}}]$ are Lipschitz (cf. (A-2)). Since the sampling is IID, and the b_i^* are centred, it follows that the right-hand side of the above display is equal to

$$n^{-2}\left\{nE[(b_1^*)^4] + 3n(n-1)E[(b_1^*)^2]^2\right\} \leq 64|x-y|^4.$$

Similarly, for $d = 3$,

$$\begin{aligned} E & \left[|\mathbb{Z}_n(x) - \mathbb{Z}_n(y)|^6\right] \\ &= \;\; n^{-3} \sum_{i,j,k,l,p,t=1}^{n} E\left[b_i^* b_j^* b_k^* b_l^* b_p^* b_t^*\right] \\ &= \;\; n^{-3}\left\{nE[(b_1^*)^6] + 3n(n-1)\left(E[(b_1^*)^3]^2 + E[(b_1^*)^2]E[(b_1^*)^4]\right)\right. \\ &\qquad\qquad \left. + \, 90n(n-1)(n-2)E[(b_1^*)^2]^3\right\} \\ &\leq \;\; 97 \cdot 2^6 \cdot |x-y|^6. \end{aligned}$$

In general, the expansion becomes

$$n^{-d}\left\{nE[(b_1^*)^{2d}] + \ldots + \binom{2d}{2\;2\;\ldots\;2} n(n-1)\ldots(n-d+1)E[(b_1^*)^2]^d\right\},$$

which is bounded above by $C(d)|x-y|^{2d}$, for some constant $C(d)$. ☐

Proof of Theorem 4.6. We first note that since $b_{\boldsymbol{A}}(x)$ is almost surely Lipschitz, then $E[b_{\boldsymbol{A}}(x_0)^2] < \infty$ for some $x_0 \in \mathcal{D}$, implies that $E[b_{\boldsymbol{A}}(x)^2] < \infty$ for all $x \in \mathcal{D}$. Therefore, convergence in finite dimensional distributions is immediate by the multidimensional central limit theorem, and it remains to prove that the process $\mathbb{Z}_n$ is tight in the space of continuous functions on $\mathcal{D}$. However, this is straightforward if we use Theorem 6.3.

The first condition with $\gamma = 2d$ and $\alpha_i = 2d$ for all i, follows immediately from Lemma 6.4 by Jensen's inequality. Thus, for the second condition we need to bound $E[\mathbb{Z}_n(x)^{2d}]$ uniformly. This follows easily since, for some fixed $x_0 \in \mathcal{D}$,

$$E\left[\mathbb{Z}_n(x)^{2d}\right] \leq C'\left(E\left[\mathbb{Z}_n(x_0)^{2d}\right] + E\left[|\mathbb{Z}_n(x) - \mathbb{Z}_n(y)|^{2d}\right]\right)$$

for come constant C' (depending on d), again applying Jensen's inequality. We have already placed a bound on the second term of the right-hand side of the above equation, and a bound on the first term follows from the central limit theorem. $\square$

Let $\mathcal{D}$ be a compact subset of $\mathbb{R}^d$. We recall a theorem of [Win]. Proposition 4.7 follows immediately.

Theorem 6.5 (SATZ 6 on page 837 of [Win])**.** *Let $\{Y(x), x \in \mathcal{D} \subset \mathbb{R}^d\}$ be a Gaussian random field such that for $\tau \to 0$ the inequality*

$$E\left[|Y(x+\tau) - Y(x)|^2\right] \leq C|\tau|^\varepsilon$$

holds for some $\varepsilon > 0$ and $0 < C < \infty$. Then for almost all realizations there exists a random number $\delta(\omega)$ so that for any $x_1, xt_2 \in \mathcal{D}$ with $|x_1 - x_2| < \delta(\omega)$ and $0 < \eta < \varepsilon/2$ the inequality

$$|Y(x_1) - Y(x_2)| \leq C_0|x_1 - x_2|^\eta$$

holds. In particular, it follows that $\{Y(x), x \in \mathcal{D}\}$ is continuous with probability one.

Proof of Proposition 4.7. To prove this result we again recall that both $b_{\boldsymbol{A}}(x)$ and $E[b_{\boldsymbol{A}}(x)]$ are Lipschitz and satisfy inequality (A-2). Therefore,

$$\begin{aligned} \operatorname{var}(\mathbb{Z}(x) - \mathbb{Z}(y)) &= \operatorname{var}(b_{\boldsymbol{A}}(x) - b_{\boldsymbol{A}}(y)) \\ &\leq E[(b_{\boldsymbol{A}}(x) - b_{\boldsymbol{A}}(y))^2] \leq |x-y|^2. \end{aligned}$$

A similar approach shows the bound for the covariance. We may now use this result, along with Theorem 6.5 to prove that the sample paths of $\mathbb{Z}$ are continuous almost surely. $\square$

Acknowledgements

The first author would like to thank Tom Salisbury for several useful discussions.

References

[BGKW] H. Breu, J. Gil, D. Kirkpatrick, and M. Werman. Linear time euclidean distance transform algorithms. *IEEE Trans. Pattern Anal. Mach. Intell.*, 17:529–533, 1995.

[Bil] Patrick Billingsley. *Convergence of probability measures.* John Wiley & Sons Inc., New York, 1968.

[BM] Adrian Baddeley and Ilya Molchanov. Averaging of random sets based on their distance functions. *J. Math. Imaging Vision*, 8(1):79–92, 1998.

[BT] B. Nebiyou Bekele and Peter F. Thall. Dose-finding based on multiple toxicities in a soft tissue sarcoma trial. *J. Amer. Statist. Assoc.*, 99(465):26–35, 2004.

[CL] Harald Cramér and M. R. Leadbetter. *Stationary and related stochastic processes. Sample function properties and their applications.* John Wiley & Sons Inc., New York, 1967.

[Cre] Noel Cressie. A central limit theorem for random sets. *Z. Wahrsch. Verw. Gebiete*, 49(1):37–47, 1979.

[DWW+] J. Ding, R.M.L. Warren, I. Warsi, N. Day, D. Thompson, M. Brady, C. Tromans, R. Highnam, and D. Easton. Evaluating the effectiveness of using standard mammogram form to predict breast cancer risk: Case-control study. *Cancer Epidemiology Biomarkers and Prevention*, 17:1074–1081, 2008.

[DZ1] M. C. Delfour and J.-P. Zolésio. *Shapes and geometries*, volume 4 of *Advances in Design and Control*. Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA, 2001. Analysis, differential calculus, and optimization.

[DZ2] Michel C. Delfour and Jean-Paul Zolésio. Shape analysis via oriented distance functions. *J. Funct. Anal.*, 123(1):129–201, 1994.

[FBF] J.H. Freidman, J.L. Bentley, and R.A. Finkel. An algorithm for finding best matches in logrithmic expected time. *ACM Trans. Pattern Anal. Softw.*, 3:209–226, 1977.

[KT] A. P. Korostelëv and A. B. Tsybakov. *Minimax theory of image reconstruction*, volume 82 of *Lecture Notes in Statistics*. Springer-Verlag, New York, 1993.

[Kun] Hiroshi Kunita. *Stochastic flows and stochastic differential equations*, volume 24 of *Cambridge Studies in Advanced Mathematics*. Cambridge University Press, Cambridge, 1990.

[Mat] G. Matheron. *Random sets and integral geometry*. John Wiley & Sons, New York-London-Sydney, 1975. With a foreword by Geoffrey S. Watson, Wiley Series in Probability and Mathematical Statistics.

[Mol] Ilya S. Molchanov. A limit theorem for solutions of inequalities. *Scandinavian Journal of Statistics*, 25:235–242, 1998.

[Nol] D. Nolan. The excess-mass ellipsoid. *J. Multivariate Anal.*, 39(2):348–371, 1991.

[Pol] Wolfgang Polonik. Measuring mass concentrations and estimating density contour clusters—an excess mass approach. *Ann. Statist.*, 23(3):855–881, 1995.

[R] R Development Core Team. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria, 2008. ISBN 3-900051-07-0.

[RP] A. Rosenfeld and J. Pfaltz. Sequential operations in digital picture processing. *J. ACM*, 13:471–494, 1966.

[SB] Larissa Stanberry and Julian Besag. Boundary reconstruction in binary images using splines. In preparation., 2008.

[SJ] Larissa Stanberry and Hanna Jankowski. Expectations of random sets and their boundaries using oriented distance functions. In preparation., 2008.

[SMW+] D. Scutt, J. Manning, G. Whitehouse, S. Leinster, and C. Massey. The relationship between breast asymmetry, breast size and the occurrence of breast cancer. *Br. J. Radiol*, 70((838)):1017–1021, 1997.

[SVW] G. Salinetti, W. Vervaat, and R. J.-B. Wets. On the convergence in probability of random sets (measurable multifunctions). *Math. Oper. Res.*, 11(3):420–422, 1986.

[SW] Gabriella Salinetti and Roger J.-B. Wets. On the convergence in distribution of measurable multifunctions (random sets), normal integrands, stochastic processes and stochastic infima. *Math. Oper. Res.*, 11(3):385–419, 1986.

[TWS] Peter F. Thall, Leiko H. Wooten, and Elizabeth J. Shpall. A geometric approach to comparing treatments for rapidly fatal diseases. *Biometrics*, 62(1):193–201, 318–319, 2006.

[Var] S. R. S. Varadhan. *Stochastic processes*, volume 16 of *Courant Lecture Notes in Mathematics*. Courant Institute of Mathematical Sciences, New York, 2007.

[Win] Wolfgang Winkler. Stetigkeitseigenschaften der Realisierungen Gauss'scher zufälliger Felder. In *Trans. Third Prague Conf. Information Theory, Statist. Decision Functions, Random Processes (Liblice, 1962)*, pages 831–839. Publ. House Czech. Acad. Sci., Prague, 1964.

Department of Mathematics and Statistics, York University
e-mail: hkj@mathstat.yorku.ca

University of Bristol, School of Mathematics
e-mail: larissa.stanberry@bristol.ac.uk